2016

Comparing the regio- and enantioselectivity of inverting versus retaining epoxide hydrolases

Mark A. Aliwalas
Wilfrid Laurier University, aliw5660@mylaurier.ca

COMPARING THE REGIO- AND ENANTIOSELECTIVITY OF INVERTING VERSUS RETAINING EPOXIDE HYDROLASES

By

Mark Anthony Aliwalas

Hon. B. Sc. Biochemistry & Biotechnology with Thesis, Wilfrid Laurier University, 2014

THESIS

Submitted to the Department of Chemistry and Biochemistry

Faculty of Science

In partial fulfillment of the requirements for the

Master of Science in Chemistry and Biochemistry

Wilfrid Laurier University

2016

Abstract

Epoxide hydrolases (EHs) are enzymes that catalyze the ring opening of an epoxide, yielding a vicinal diol. This exciting class of enzymes is often associated with natural product and small molecule biosynthesis. One interesting class of natural products that epoxide hydrolases are involved in the biosynthesis of is the enediynes. Previously, 5 enediyne-associated epoxide hydrolases have been characterized, revealing an inverting-versus-retaining paradigm for enediyne epoxide hydrolases. The work described herein sets the table for future studies involving the probing of the α/β-epoxide hydrolase mechanism for styrene oxide. Kinetic and regioselectivity characterization of the enediyne-associated CynF and SghF has determined that SghF may be a robust starting platform for engineering efforts in the hopes of generating enantiomerically pure diols through biocatalysis. SghF represents the most efficient (K_{cat}/K_m of 190 for (*S*)-styrene oxide) and enantioselective (E=655, in favour of the (*S*)-epoxide) enediyne-associated epoxide hydrolase characterized to date. The identification of conserved residues at positions Y138, L187, A188, D189, P190, E191, H192, A194, F229, Q303, L304, and Q370 may represent critical sites for investigation in the hopes of generating a robust catalytic toolkit for the production of enantiomerically pure diols from a wide variety of epoxide hydrolases and substrates. Finally, regioselectivity characterization of CynF and SghF has confirmed that CynF is a retaining epoxide hydrolase, and SghF is an inverting epoxide hydrolase, thereby expanding the inverting versus retaining paradigm for enediyne epoxide hydrolases.

Acknowledgements

I would like to extend my sincerest gratitude to Dr. Geoff Horsman and the members of the Horsman lab throughout the years: Maddison Bibby, Matt Chi, Jayne Kelso, Kissa Batul, Navjit Singh, Rebecca Sullivan, Ali Al-Zein, Warad Al-Far, Alena Pratasouskaya, Julia Szusz, and Zahra Yussuf. I would also like to thank my cats, Monty and Millie. I would not have made it past all of the garbage that life has thrown at me during this degree if it weren't for all of you. Secondly, I would like to thank Dr. Michael Suits for help with protein crystallography, and Joe Meissner and Joel Weadge for technical advice and use of equipment. Finally, I would also like to pay respects to those who were unable to witness the completion of my Master's degree: my maternal grandmother, Angelita De Guzman, my paternal grandmother, Regina Aliwalas, and most importantly my mother, Christy Aliwalas. I love you all and miss you dearly. Wish me luck.

Author's Note

The CytoScape network used in these works was generated by Danielle Kochen. Cloning and initial crystal screening for SghF was performed by Navjit Singh, and refinement of the crystal growth conditions and crystal structure resolution was performed by Jonah Nechacov. These works are included in this thesis due to their relevance to the subject matter; I take no credit for work done by others.

Table of Contents

Tables and Figures

Introduction

Epoxide Hydrolases

Epoxide hydrolases (EHs) are enzymes that catalyze the ring opening of epoxides, yielding a corresponding diol. Putative EHs have been found in approximately 20% of all sequenced organisms [1], and the majority of EHs that have sequence information available are contained within one of two structural families: the limonene-1,2-epoxide hydrolase fold family, and the α/β hydrolase fold family[2]. The majority of EHs are contained within the α/β hydrolase fold family[2]. Enzymes contained within this family possess a main domain comprised of a single central β-sheet that is enclosed by several α-helices and a variable cap domain positioned at the top of the substrate binding site [3]. Epoxide ring opening by α/β EH catalysis is achieved through a strongly conserved Asp/Ser/Cys-His-Asp/Glu catalytic triad, located on the main domain which highly resembles other α/β-fold enzymes[2]. The substrate epoxide is anchored by two conserved tyrosine residues located on the cap region of the enzyme [4]. The nucleophile for α/β hydrolase fold EHs is an invariable aspartate [1] which attacks at the least hindered carbon of the epoxide ring, leading to the formation of an alkyl enzyme intermediate (**Figure 1**) [5]. An Asp-His-Asp triad then activates a water molecule that hydrolyzes the ester, releasing the product diol [6].

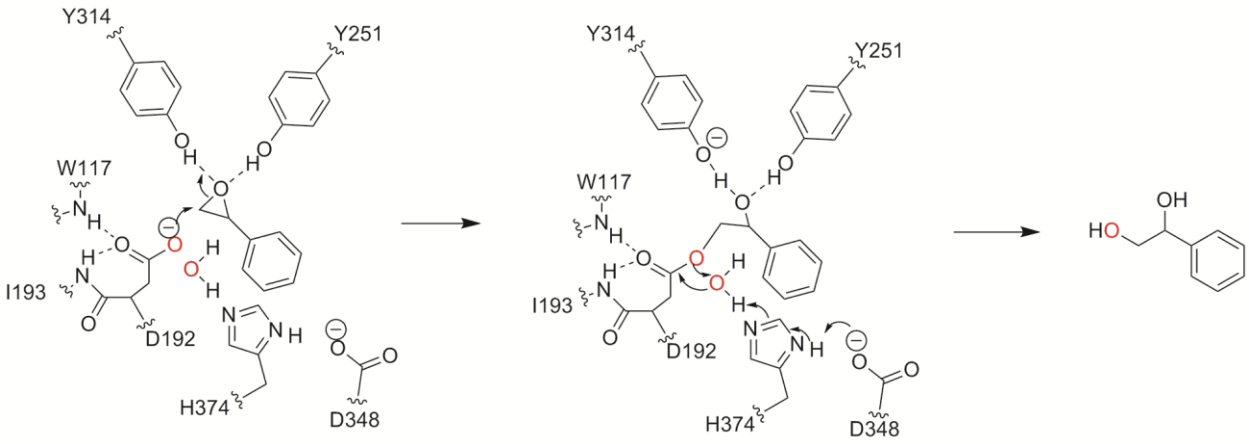


Figure 1. Proposed catalytic mechanism for the hydrolysis of styrene oxide by an α/β-fold epoxide hydrolase. The active site residues reflect the predicted active site residues in the AnEH from *Aspergillus niger*

EHs are important for the biosynthesis of some microbial natural products that possess a wide array of therapeutically relevant bioactivities, such as antimicrobial, antifungal, and antitumour properties [7]. While the scaffold of many of these clinically relevant secondary metabolites are synthesized by type I polyketide synthases (PKSs) and non-ribosomal peptide synthases (NRPSs), EHs are among several tailoring enzymes that are involved in the decoration of natural product scaffolds to yield the mature molecule [8,9].

One of the important classes of microbial natural products are the enediynes. EHs are important in the biosynthesis of enediynes that possess a 9-membered ring structure, or enedidyne core, which acts as the "warhead" that induces breaks in double-stranded DNA [10]. Enediynes have great potential as antitumour antibiotics, with extreme potencies (IC_{50}s as low as 10^{-3}pM for selected cancer cell lines). The biosynthesis of the enediyne core is poorly understood; however recent work has provided understanding on how special peripheral moieties are added to 9-membered enediyne cores via a vicinal diol handle [11]. Interestingly, it appears as though the vicinal diol is generated as a result of the hydrolysis of an (*S*)-epoxide by an EH [12]. Furthermore, the resultant configuration of the stereocenter in the mature enediyne is dictated by the regioselectivity in the associated EH (**Figure 2**). For instance, enediynes possessing an (*R*)-vicinal diol are generated from the hydrolysis of the (*S*)-epoxide precursor catalyzed by an EH that inverts the stereochemistry of the precursor by attacking at the more hindered carbon [13]. In contrast, (S)-vicinal diols are generated from the hydrolysis of the (*S*)-epoxide precursor by attacking at the less hindered carbon. EHs that invert the stereochemistry of the epoxide precursor are called "inverting" EHs, while EHs that retain the stereochemistry of the epoxide precursor are called "retaining" EHs [13].

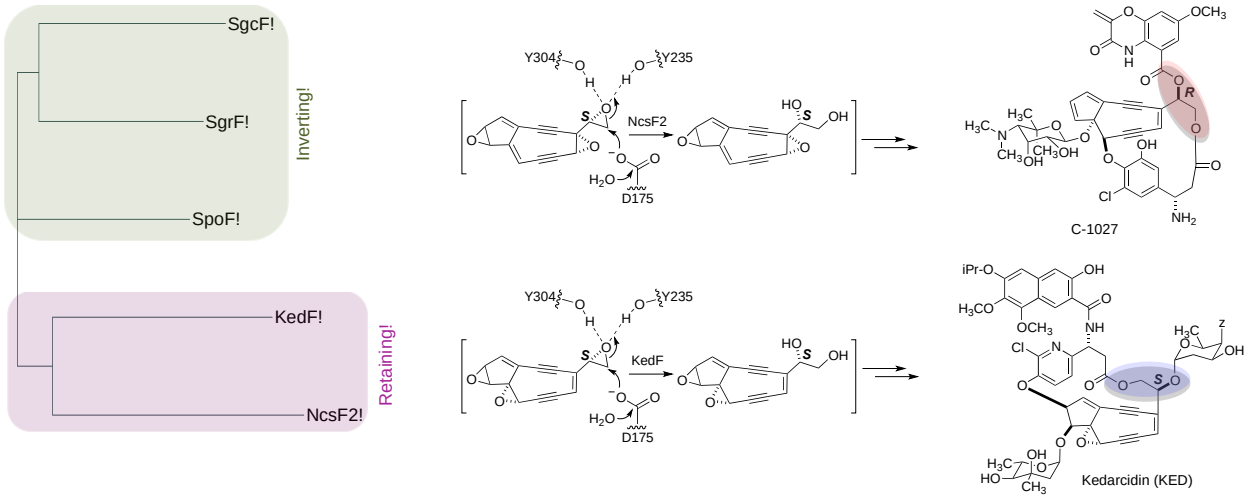


Figure 2. Familial classification of currently characterized enediyne epoxide hydrolases. Adapted from [13].

To date, 5 enediyne-associated EHs have been characterized and reveal some unusual features compared to canonical EHs. A sequence alignment of these enzymes reveals features typical of EHs: (i) the catalytic triad residues D175, D336, and H363 (KedF numbering) and (ii) the two epoxide-binding residues at positions Y236/W236 and Y304 [13]. Surprisingly, although all characterized EHs possess Tyr residues that bind to the epoxide oxygen, the three known inverting EHs have had one of the canonical tyrosine residues replaced with tryptophan at position 236. Previous phylogenetic analysis suggests that there is a familial link for inverting versus retaining EHs [13].

Earlier research done by Horsman *et al.* on the inverting EH SgcF [13] has probed the molecular determinants of EH catalysis through the mutation of the active-site tryptophan (W236) to the canonical tyrosine. Indeed, the mutation did marginally increase the retaining activity of the enzyme; however, only with an accompanying mutation in the neighboring residue (W236Y/Q237M) was there an appreciable increase in the enzyme's retaining activity towards styrene oxide [13]. Despite this, there was still significant inverting activity in the double mutant. Clearly, epoxide hydrolase regioselectivity is far more complex than the diagnostic tyrosine to tryptophan mutation, and the molecular determinants of

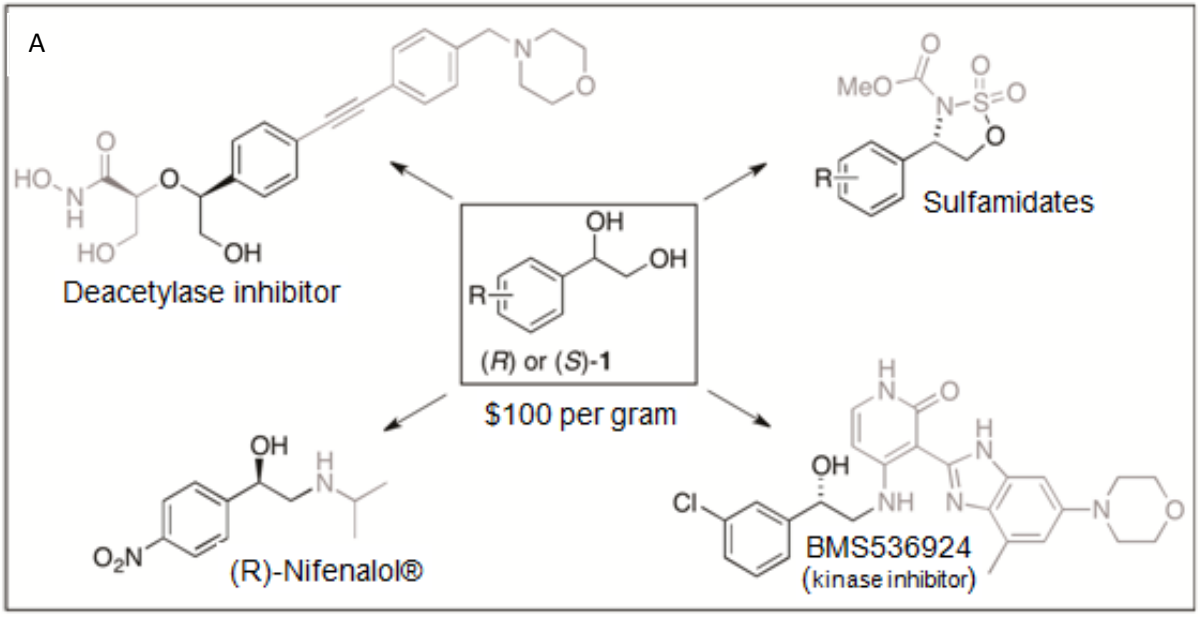
EH inverting versus retaining catalysis remain elusive. Comprehension of these molecular determinants is vital to unlocking the potential for EHs to be used in biocatalysis.

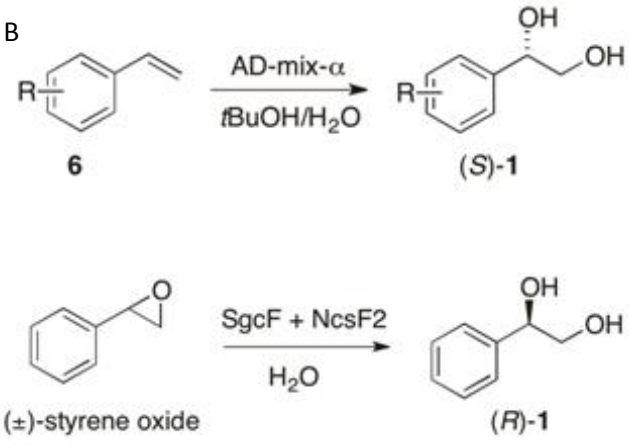
Biocatalysis

Biocatalysis involves the application of both enzymes and microbes as catalysts in synthetic chemistry [14]. It may also include the use of natural catalysts for processes which have not yet naturally evolved. First employed centuries ago, biocatalysis has grown in complexity from the use of plant and microbial cell extracts to perform simple chemical transformations to the industrial preparation of high value chirally pure synthetic intermediates for pharmaceuticals [15]. The discovery of microbial epoxide hydrolases has broadened the scope of EH research to include their application in biocatalysis[1]. EHs are prime candidates for biocatalysis due to the wide range of substrates that they are capable of accepting, ranging from small aliphatic to bulky polyaromatic epoxides[2]. Furthermore, microbial epoxide hydrolases do not require expensive cofactors, prosthetic groups or metal ions for activity, making them attractive for large-scale industrial applications[16]. Indeed, not only are microbial epoxide hydrolases easy to produce in large quantities, but the cloning and overexpression of several enantioselective epoxide hydrolases has facilitated their large-scale production and improvement of their biocatalytic capabilities by either random or site-directed mutagenesis [1,6,16,17]. Furthermore, powerful modeling tools and algorithms have allowed researchers to engineer biocatalysts to perform in unnatural solvents, increased temperature stability, and increased enantio- and regioselectivity [18–23]. Additionally, The establishment of new techniques for creating more efficient screening libraries such as triple codon saturation mutagenesis (TCSM) and fast screening techniques such as on-chip analysis of enantiomeric transformations continues to further the potential practicality and applicability of EHs in biotransformations [23,24].

Despite the alluring nature of the use of enzymes for biocatalytic transformations, there still remain several drawbacks to enzyme catalysis that must be overcome before their widespread use in industry is realized. A major drawback to the use of enzymes for industrial scale biotransformations is often their instability to the conditions employed. Although researchers are able to engineer and improve upon natural catalysts, the final engineered enzyme is frequently unsuitable for use in industrial processes. The conditions employed are often significantly different when compared to the enzymes' natural environment in the cell, including non-natural substrates, non-aqueous conditions, and extremes of pH [25]. A more robust understanding of EH catalysis and engineering will inform the design and engineering of EHs that are capable of generating pure stereoisomers from a wide variety of substrates in a predictable manner.

Regioselective EHs could be used as biocatalysts to generate enantiomerically pure (*R*)- or (*S*)-**1**, which can be used as a chiral substrate in the production of high value synthetic intermediates such as sulfamidates and bioactive compounds such as the kinase inhibitor BMS536924, the β-blocker (*R*)-Nifenalol ®, and deacetylase inhibitor [21,26–28]. Furthermore, it has been shown that EHs are not only comparable in yield and efficiency to conventional catalysts, they can often exceed the capabilities of conventional organic synthesis (**Figure 3**) [12,21]. Moreover, dependence on heavy-metal based catalysts presents an environmental risk.





97% ee and > 80% yield
AD-mix is heavy metal based
$5.60/g of catalyst
$25/g for 2-bromostyrene
$8.50/g for 4-bromostyrene

99%ee and 87% yield
$1.10/g for each catalyst
$0.30/g for styrene oxide

Figure 3. (A) Enantiomerically pure 1-phenyl-1,2-ethanediol as a chiral substrate for high value pharmaceuticals. (B) Comparison of a conventional catalyst against the combined action of the EHs SgcF and NcsF2 [13]

Bioinformatics and Computational Tools for Annotating Enzyme Function

The amount of protein sequence data available has increased exponentially in the last decade; the number of sequence entries in the UniProtKB database exceeds 65,000,000 as of early July 2016. The bounty of protein sequences available is a boon for researchers, as the understanding of enzyme functions and their associated metabolic pathways should allow for advances in not only industry, but in

medicine, chemistry, and synthetic biology [29]. However, the number of enzymes that have been assigned reliable *in vitro* functions is greatly outnumbered by the genes that have only been assigned putative function. These genes only share a marginal amount of sequence similarity to related enzymes, making assumptions on their molecular function difficult. Furthermore, a large amount of functional annotations assigned to genes are incorrect; as much as 40% of genes in public databases are misannotated [30].

As such, there is an increasing need to determine reliable functions for biochemically uncharacterized proteins. Given the number of unknown sequences, biochemical experimentation alone is not a feasible strategy [29]. Computational tools can be used to complement biochemical experimentation and reliably annotate functions for proteins that cannot be characterized experimentally. Bioinformatic analyses can assign tentative functions by grouping similar enzymes into clusters with a single hypothesized function. Additionally, biochemical characterization of two enzymes with opposing activities (i.e. inverting and retaining EHs) that are closely related allows for the in-depth analysis of subtle differences in enzyme sequence, structure and function that leads to opposing regioselectivity in EHs.

The CytoScape network generated through the Enzyme Function Initiative [29] shown in **Figure 4** demonstrates the relationship of known EHs to each other in 2-dimensional space. Closer examination of this network reveals that the EHs containing the tryptophan mutation that is diagnostic of inverting catalysis are all closely related to one another. Furthermore, this cluster of putative inverting EHs is surrounded by EHs which have both of the conventional epoxide-binding tyrosines, which are predicted to be retaining EHs. Two highly related EHs have been selected for investigation. The EHs SghF (from *Streptomyces ghanaensis* sp. ATCC-14672) and CynF (from *Streptomyces* sp. CNT-179) have been selected because: i) they are highly related and found in gene clusters encoding for enediynes and ii)

they are predicted, by the familial classification of enediyne epoxide hydrolases, to possess opposing regioselectivities.

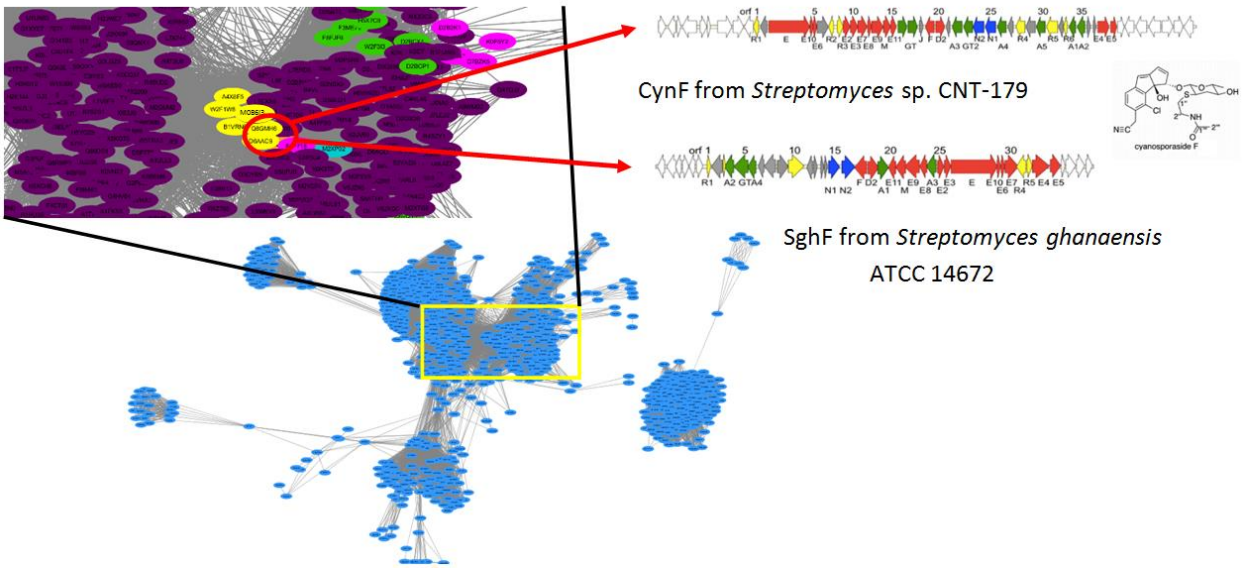


Figure 4. Full CytoScape cluster for epoxide hydrolases. Each node in the cluster is representative of a protein sequence. A line connecting two nodes is indicative of relatedness. Inset clustering of tryptophan epoxide hydrolases, and the genes selected for this study. Purple nodes represent EH sequences which possess the canonical tyrosine residue at position, green nodes respresent EHs which possess a phenylalanine residue, pink nodes represent EHs which possess a histidine residue, and yellow nodes represent EH sequences that possess a tryptophan residue.

Objectives

Bioinformatic analysis has already identified two epoxide hydrolases, CynF and SghF, which are highly similar, but are expected to exhibit opposing regioselectivities. Biochemical and structural characterization of these enzymes will lead to a more robust understanding of EH catalysis and selectivity. It is also important to be able to identify suitable EHs for the purposes of biocatalysis. Therefore, the goals of this project were as follows:

(1) Perform a bioinformatics analysis on the known tryptophan EHs and compare them to similar tyrosine EHs that are nearby in the SSN. This may lead to identification of the conserved

catalytic residues within CynF and SghF. This may also lead to novel amino acid sites to further investigate the molecular determinants of EH regio- and enantioselectivity.

(2) Biochemically characterize and solve the protein crystal structure of the EHs CynF and SghF. Characteristics of these enzymes may be used to assign reliable activities to nearby enzymes within the EH CytoScape network and further expand the inverting versus retaining paradigm of enediyne epoxide hydrolases.

(3) Generate epoxide hydrolase mutants that are expected to alter or hinder the EH catalytic mechanism. Structural characterization of these mutants would allow visualization and probing of each step of EH catalysis.

Materials

Instrumentation

Agarose gel electrophoresis was performed using a Midi 10 Gel Electrophoresis system (VWR, Radnor, PA). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using a Mini PROTEAN Tetra Gel Electrophoresis chamber (Bio-Rad, Berkeley, CA). Power for electrophoresis was provided by a 300V power source (VWR, Radnor, PA). Ultrapure water was obtained using a Synthesis A10 Water Purification System (Milli-Q, Etobicoke, ON). Microcentrifugations at room temperature were performed using a 5415C microcentrifuge (Eppendorf, Hamburg, Germany). Centrifugations done at 4 ^{o}C were performed using a Heraeus Multifuge X1R (ThermoFisher Scientific, Waltham, MA). High performance liquid chromatography was done using a Shimadzu Prominence HPLC (Mandel, Guelph, ON). Reverse-phase HPLC was done using an Alltima C18 column ((5µm, 150mm x 4.6 mm), Alltech, Columbia, MD). Chiral HPLC was done using a Chiralcel OD-H column ((5µm, 4.6mm x 250mm) Diacel, Westchester, PA). Miscellaneous spectrophotometry (Bradford Assay and optical density analysis) was performed using a Spectramax 384 Plus (Molecular Devices, Sunnyvale, CA).

Shaking incubation of cell cultures was done using a MaxQ incubating shaker (ThermoFisher Scientific, Waltham, MA). Incubation of LB agarose plates was done using a Heratherm Compact microbiological incubator (ThermoFisher, Waltham, MA). Vortexing was done using a Thermolyne MaxiMix II vortexer (ThermoFisher, Waltham, MA). Sonic disruption of cells was done using a Q125 sonic processor (Q-Sonica, Newtown, CT). Ultrafiltration was done using a Macrosep Advance 30k MWCO Centrifugal Device (Pall Corporation, Mississauga, ON). Buffer exchange and desalting was done using EconoPac 10DG columns (Bio-Rad, Berkeley, CA). Anion exchange was performed using a Hi-Trap Q FF 5mL Anion Exchange column (GE Life Sciences, Mississauga, ON). Size exclusion chromatography was performed using an Enrich SEC 350 24mL gel filtration column (Bio-Rad, Berkeley, CA). Fast protein liquid chromatography was done using a NGC Quest Medium Pressure Chromatography system (Bio-Rad, Berkeley, CA). Epoxide hydrolase kinetics was monitored spectrophotometrically using a Cary 60 UV/Visible spectrophotometer with an attached PCB1500 water peltier system (Agilent Technologies, Santa Clara, CA) to maintain a constant temperature. Initial protein crystallography screens were performed using a Gryphon crystallography robot (Art Robbins Instruments, Sunnyvale, CA), and microscopy of potential crystallography hits was done using SZX16 microscope with an attached Highlight 3100 light source (Olympus, Toyko, Japan).

Kits

Extraction of DNA from agarose gels and general DNA purification was performed using the Wizard SV Gel Extraction and PCR Clean-Up kit (Promega, Madison, WI). SDS-PAGE gels were cast using the TGX FastCast Stain-Free Acrylamide kit (Bio-Rad, Berkeley, CA). Plasmid DNA minipreps were done using a PureYield Plasmid miniprep kit (Promega, Madison, WI). High fidelity PCR was done using the Phusion DNA high-fidelity DNA polymerase (ThermoFisher Scientific (Waltham, MA). Adenine overhang addition was done using EconoTaq polymerase (Lucigen, Middleton, WI). Cloning was done using the pGEM-T-Easy vector system (Promega, Madison, WI). Initial crystal screening was done using Index and

Crystal Screen 1 and 2 (Hampton Research, Aliso Viejo, CA) and MCSG-1, -2, -3, and -4 (Anatrace, Maumee, OH).

Chemicals and Other Materials

The following is a list of chemicals and suppliers used in the methodology of this project: pUC57-SghF (BioBasic, Markham, ON); pET28 (Geoff Horsman, Madison, WI); pET29 (Geoff Horsman, Madison WI); Cosmid O19 containing the cynF gene (Amy Lane, Jacksonville, FL); Tryptone (ThermoFisher, Waltham, MA); yeast extract (Amresco, Cleveland, OH); NaCl (Amresco, Cleveland, OH); polyethyleneglycol 8000 (ThermoFisher Scientific, Waltham, MA); bacteriological agar (Alfa Aesar, Haverhill, MA); Sterile DMSO (ThermoFisher Scientific, Waltham MA); non-sterile DMSO (BioBasic, Markham, ON); $MgCl_2 \bullet 6H_2O$ (Amresco, Cleveland, OH); Agarose (Amresco, Cleveland, OH); Sucrose (VWR, Radnor, PA); Bromophenol Blue (Amresco, Cleveland, OH); dNTPs (10mM each) (ThermoFisher Scientific, Waltham, MA); 50mM $MgCl_2$ (ThermoFisher Scientific, Waltham, MA); Ethanol (Commercial Alcohols, Brampton, ON); Coommassie Blue G-250 (Amresco, Cleveland, OH); Phosphoric Acid (Anachemia, Missisauga, ON); Restriction enzymes (ThermoFisher Scientific, Waltham, MA); Egg white lysozyme (BioBasic, Markham, ON); IPTG (BioBasic, Markham, ON); His-Pur Ni-NTA slurry (ThermoFisher Scientific, Waltham, MA); DNase I (Sigma Aldrich, St. Louis, MO); Ethidium bromide (Millipore, Etobicoke, ON); Ampicillin sodium salt (VWR, Radnor, PA); Kanamycin sulfate (Amresco, Cleveland, OH); X-GAL (TEKNOVA, Hollister, CA); NaOH (Amresco, Cleveland, OH); Racemic styrene oxide (Sigma Aldrich, St. Louis, MO); Ethyl acetate (VWR, Radnor, PA); Acetonitrile, HPLC grade (ThermoFisher Scientific, Waltham, PA); Isopropanol, HPLC grade (ThermoFisher Scientific, Waltham, PA); Hexane, HPLC Grade (ThermoFisher Scientific, Waltham, PA); (*R*)-styrene oxide (Sigma Aldrich, St. Louis, MO); (*S*)-styrene oxide (Sigma Aldrich, St. Louis, MO); racemic 1-phenyl-1,2-ethanediol (Sigma Aldrich, St. Louis, MO); (*R*)-1-phenyl-1,2-ethanediol (Sigma Aldrich, St. Louis, MO); (*S*)-1-phenyl-1,2-ethanediol (Sigma Aldrich, St. Louis, MO); Sodium periodate (Sigma Aldrich, St. Louis, MO); Bis-Tris (VWR, Radnor, PA); Sodium acetate

(BioShop, Burlington, ON); Hydrochloric acid (VWR, Radnor, PA); Calcium chloride (Amresco, Cleveland, OH); Ethylene glycol (BioShop, Burlington, ON); Tris (VWR, Radnor, PA); Imidazole (ThermoFisher Scientific, Waltham, MA); Sodium phosphate monobasic (Amresco, Cleveland, OH); Sodium phosphate dibasic (Amresco, Cleveland, OH); Polyethylene glycol 3.35k (Hampton Research, Aliso Viejo, CA); Polyethylene glycol 8k (ThermoFisher Scientific); Polyethylene glycol 400 (Hampton Research, Aliso Viejo, CA); Polyethylene glycol monomethyl ether (Hampton Research, Aliso Viejo, CA); Ammonium Sulfate (VWR, Radnor, PA); Tetramethylethylenediamine (Amresco, Cleveland, OH); Carbenicillin sodium salt (ThermoFisher Scientific, Waltham, MA); Ammonium persulfate (BioShop, Burlington, ON).

Methodology

All agarose gels were prepared by combining 1.0% (w/v) of agarose to the appropriate volume of Tris-Acetate-EDTA (TAE, 0.04M Tris-Acetate, 0.001M ethylenediaminetetraacetic acid (EDTA)) buffer and heating in a microwave until fully dissolved. The gel was then left to cool and ethidium bromide (0.5µg/mL of gel) was added and poured into an appropriate plastic mold. A well-forming comb was then inserted and the gel was left to solidify at room temperature. Once solidified, the gel was placed in the gel tray and submerged in TAE buffer. All unstained DNA was diluted 1:1 with DNA staining solution (0.25% bromophenol blue, 40% (w/v) sucrose) to aid in DNA sample addition into the gel wells. DNA samples were then loaded into the wells using a micropipette and the cover was placed on the chamber. The electrophoresis apparatus was then attached to a power source and run for 45 minutes at 100V. The gel was then removed from the gel tray and placed onto a Bio-Rad Gel Doc Easy UV Tray for gel imaging.

Bioinformatics Analysis

The CytoScape network generated was used to identify the cluster of EHs which possess the tryptophan mutation. The 6 nearest EHs possessing the canonical tyrosine residue in the SSN were

selected for analysis by sequence alignment and aligned with the 6 identified tryptophan-containing EHs. The amino acid sequences of the selected genes were retrieved from the UniProt KB database [31] and used for sequence alignment by Clustal Omega [32,33].

Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was performed using a Bio-Rad C1000 Touch Thermal Cycler. A 100µL master mix was made by combining 55µL of nuclease-free water, 20µL 5X HF Phusion Green Buffer, 10µL dimethyl sulfoxide (DMSO), 3µL deoxynucleotide phosphate (dNTP) mix, 2µL 50mM $MgCl_2$, 3µL forward primer, 3µL reverse primer, 3µL forward primer, 3µL template DNA, and 1µL Phusion HiFi polymerase. The master mix was then dispensed into 10uL aliquots for thermal cycling. Cycling parameters are outlined in **Table 1.**

Table 1. Cycling parameters for PCR amplification

Step #	Temperature (^{o}C)	Time (min:sec)
1	**94**	**2:00**
2	**94**	**0:15**
3	**Gradient 50-65**	**0:10**
4	**72**	**1:10, return to step 2 x30**
5	**72**	**5:00**
6	**4**	**Infitinite hold**

Preparation of Luria Broth and Luria Broth Agar for *Escherichia coli* Cultivation

Luria Broth (LB) was prepared by combining 10g/L tryptone, 5g/L yeast extract, and 10g/L NaCl in ultrapure water from a Milli-Q Synthesis A10 Water Purification system from Millipore (Etobicoke, ON). The prepared LB was then sterilized using the liquid sterilization program in a Steris Amsco Lab 250

steam sterilizer (Mississauga, ON). LB Agar was prepared by adding an additional 1g/L of agar to LB prior to sterilization.

Preparation of *Escherichia coli* Competent Cells

A glycerol freezer stock of *E.coli* DH5α or BL21 was streaked onto Lysogeny broth (LB) agar and allowed to grow overnight at 37 ^{o}C. A single colony was then picked and used to inoculate 50mL of LB and allowed to grow at 37 ^{o}C with shaking until the optical density at 600nm (OD_{600}) reached approximately 0.5. The culture was transferred to 2 ice-cold sterile 50mL conical tubes and let to sit on ice for 10 min. The cultures were then centrifuged at 4 ^{o}C for 10 min at 2700 x *g*. The supernatant was then poured off and cells were resuspended in 10mL of ice cold sterile TSS buffer (2.0% (w/v) NaCl, 2.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 10% (w/v) polyethylene glycol 8000, 0.5% (v/v) dimethyl sulfoxide, 1.0% (w/v) $MgCl_2 \bullet 6H_2O$, pH 6.5), dispensed into 200µL aliquots and flash frozen using liquid nitrogen. The competent cells were then stored in a Revco Elite upright freezer at -80 ^{o}C until needed.

Transformation of DNA into *E. coli* Competent Cells

Competent cells were allowed to thaw on ice, and approximately 30-80ng of DNA were pipetted directly into the cells and allowed to sit on ice for 20-30 min. The cells were then heated at 42 ^{o}C for 2 min. After heating, 1mL of LB was added and the cultures were allowed to incubate at 37 ^{o}C for 1h with agitation. The cultures were then spun down at full speed for 1 minute, and most of the supernatant was discarded, while approximately 100µL of supernatant was reserved. The cell pellets were resuspended in the reserved supernatant and plated onto LB agar containing 10µg/mL kanamycin for pET29 constructs and 20µg/mL ampicillin, 80µg/mL X-Gal and 200µg/mL IPTG for pGEM constructs. The plates were then incubated for at least 12 hours at 37 ^{o}C. Positive transformants for incorportation into pET29 are indicated by growth, while positive transformants for pGEM are indicated by growth and the absence of a blue colour change. Colonies were then picked into 3mL of LB containing the appropriate

antibiotic using a sterile inoculating loop and allowed to grow overnight at 37 ^{o}C with shaking. The plasmid DNA is then purified through plasmid DNA miniprep. If required, a glycerol freezer stock is first made using 20% sterile glycerol and adding cell culture with *E.coli* DH5α containing the construct to 1mL and stored at -80 ^{o}C.

Preparation of DNA Primer Stocks

DNA primers were purchased as dry residue from Integrated DNA Technologies (IDT, Coralville, IA). A 100µM stock was prepared by adding an appropriate amount of nuclease-free water and the DNA was dissolved thoroughly by agitation with a vortexer. An additional 100µL of 10µM working stock was then prepared by combining 10µL of 100µM stock with 90µL of sterile water. Unless otherwise noted, the 10µM working stock was used for polymerase chain reactions.

Removal of Gycerol and Sodium Azide from Centrifugal Device Membranes

Centrifugal devices were rinsed by flow through of 10mL of ultrapure water by centrifugation for 10 minutes at 2,700 *x g*. Both the filtrate and the retentate were disposed. A second rinse with 0.05N NaOH was performed by centrifugation for 10 minutes at 2,700 *x g*, and once again the filtrate and retentate were disposed of. A final 10 minute rinse at 2,700 *x g* with ultrapure water was then done to remove any residual NaOH. The columns are then stored in ultrapure water at room temperature until needed.

Restriction Enzyme Digests

Restriction enzyme digests were performed using up to 1µg of DNA, 2µL of 10X FastDigest Green Buffer, 1µL of an appropriate restriction enzyme (1µL of each enzyme for double digests). Nuclease-free water was used to dilute the reaction to 20µL. The digests were allowed to incubate at 37 ^{o}C for 20 minutes before loading onto a 1% agarose gel for separation. Electrophoresis was performed as

described in previous sections. Digests that were to be used downstream were extracted from the gel after electrophoresis.

pGEM-T Easy Ligation

DNA was ligated into pGEM-T-Easy using the following mixture: 60ng of insert DNA, 20ng of pGEM-T-Easy, 5µL of 2X Rapid Ligation Buffer, and 1µL of T4 DNA Ligase. Where required, nuclease-free water was used to bring the total reaction volume to 10µL. A separate positive control was prepared by replacing the insert DNA with 3µL of Control Insert DNA, and a negative control was prepared by replacing the insert DNA with nuclease-free water. The reactions were incubated overnight at 4 ^{o}C, and transformed into *E. coli* DH5α.

pET29 Ligation

DNA was ligated into pET29 using the following mixture: 20ng of pET29, 60ng of insert DNA, 1µL of 10X T4 Ligase Buffer, and 1µL of T4 DNA ligase. A negative control was performed by replacing the insert DNA with nuclease-free water. The reactions were incubated at 4 ^{o}C overnight, and transformed into either *E. coli* DH5α for plasmid miniprep, or *E. coli* BL21 (DE3) for protein expression.

Cloning of the CynF and SghF genes that encode epoxide hydrolases

Cosmid AD12 containing the CynF gene from *Streptomyces* sp. CNT-179 was obtained from Amy Lane, U. North Florida. The CynF gene was amplified using the polymerase chain reaction (PCR) with primers CynF-NdeI-F and CynF-XhoI-R (**Table 2**). A master mix was made as previously described and dispensed into 10µL aliquots for the annealing gradient. Cycling parameters used are summarized in **Table 1.** The CynF PCR product has additional 3' XhoI and 5' NdeI sites, as well as a MAH_6VD_4K N-terminal tag to allow for production of a His_6-tagged epoxide hydrolase product similar to previously characterized SgcF produced from pBS1096 [11]. The PCR product was run on a 1% (w/v) agarose gel and

amplicons corresponding to the correct size were extracted using the Wizard Gel Extraction and PCR Cleanup kit. A single adenine-overhang was then added to assist in annealing to pGEM T-Easy as previously described**.** The gel extract was then ligated into pGEM-T Easy as previously described to produce the sequencing construct CynF-pGEM. The construct CynF-pGEM was then sent to the Centre for Applied Genomics in Toronto, Ontario for sequencing via the T7 forward and SP6 reverse sequencing primers. The retrieved sequence was then confirmed using the NCBI BLAST tool.

Table 2. Primers used for cloning procedures described. Restriction enzyme cut sites are underlined. Fusion tags are bolded.

Primer Name	**Sequence**
CynF-NdeI-F	5'-**CATATGGCACATCACCACCACCATCACGTGGATGACGACGACAAG** ATGGATCCCTTTCGGATCGACATC-3'
CynF-XhoI-R	5'-**CTCGAG**CTACCGGCCGGGCAGCGATCG-3'

The vector pUC57-SghF containing the SghF gene from *S. ghanaensis* sp. ATCC-14672 was purchased from BioBasic (Markham, ON). The restriction sites for NdeI and XhoI present at the 5' and 3' end of the gene allow for the restriction enzyme digest and ligation of the SghF gene into the vector pET28 from Geoff Horsman (Madison, WI). The above method was also used to clone the CynF gene from the CynF-pGEM construct into the vector pET29 from Geoff Horsman (Madison, WI).

Preparation of EH Mutants by Overlap Extension PCR

PCR primers that flank the gene of interest were ordered from Integrated DNA Technologies (Coralville, IA) that contain a mismatched amino acid codon at an appropriate location to yield a desired mutant (**Table 3**). PCR primers matching the 5' end and 3' end of the SghF gene were also ordered from Integrated DNA Technologies. Overlap extension PCR was performed in two rounds: the first round

includes the separate PCR amplification of the sections of the genes upstream and downstream of the mutation. The CynF mutant with histidine 363 mutated to adenine (CynF-H363A) was prepared using two separate PCR reactions, utilizing CynF-pGEM as a template. The first reaction was performed using the primers CynF-NdeI-TAG-F and CynF-H363A-R to amplify the CynF gene upstream of the desired mutation. The downstream portion of the gene was then prepared using the primers CynF-H363A-F and CynF-XhoI-R and CynF-pGEM as the template sequence. After extraction via the Wizard Gel Extraction and PCR Cleanup Kit from Promega, Fragments A and B were then spliced together via PCR using both fragments as the template sequence, and the primers CynF-NdeI-TAG-F and CynF-XhoI-R. This final full length CynF mutant CynF-H363A was then ligated into pGEM-T-Easy (Promega) after addition of the adenine overhang using the techniques previously described.

Table 3. Primers used for the generation of EH mutants by overlap extension. Restriction enzyme cut sites are underlined. Fusion tags are bolded.

Primer Name	**Sequence**
CynF-NdeI-TAG-F	5’-**CATATGGCACATCACCACCACCAT** -3’
CynF-XhoI-R	5’-**CTCGAG**CTACCGGCCGGGCAGCGATCG-3’
CynF-H363A-F	5’-CGCGAGGCGGCGCGTTCGCGGCC-3’
CynF-H363A-R	5’-GGCCGCGAACGCGCCGCCTCGCG-3’
CynF-H363Q-F	5’-CGCGAGGCGGCCAGTTCGCGGCC–3’
CynF-H363Q-R	5’-GGCCGCGAACTGGCCGCCTCGCG–3’
CynF-D175N-F	5’-CGCGGGCGGGAACTGGGGATC–3’
CynF-D175N-R	5’-GATCCCCAGTTCCCGCCCGCG-3’
SghF-Nat-F	5’-**CATATG**CGTCCTTTCCGCATCGAGATC-3’
SghF-H364A-F	5’-CAAGGGTGGTGCGTTCGCCGCGATG–3’
SghF-H364A-R	5’-CATCGCGGCGAACGCACCACCCTTG–3’
SghF-H364Q-F	5’-CAAGGGTGGTCAGTTCGCCGCGATG–3’

SghF-H364Q-R	5’-CATCGCGGCGAACTGACCACCCTTG–3’
SghF-D176N-F	5’-CAGGGTGGAAACTGGGGTATGGCC-3’
SghF-D176N-R	5’-GGCCATACCCCAGTTTCCACCCTG-3’

The above method was used to generate the EH mutant constructs CynF H363A, CynF H363Q, CynF D175N, SghF H364A, SghF H364Q, and SghF D176N. The pGEM-T-Easy ligated mutant epoxide hydrolase constructs were then sent for sequencing at The Center for Applied Genomics, and sequence verification as well as confirmation of successful mutation was done using the NCBI BLAST tool. The SghF and CynF mutants were then subcloned into pET28 and pET29, respectively using the methods described above for protein overexpression and purification.

Overexpression and Purification of EHs

Transformation of the expression constructs pET28-SghF and pET29-CynF into *E. coli* BL21 and overproduction and purification of the resultant EHs were performed as previously established and described above. A single transformant was picked and used to inoculate 50mL of LB containing 10μg/mL kanamycin and allowed to grow overnight at 37 ^{o}C with agitation. 4L of LB with 10μg/mL of kanamycin was inoculated with the overnight growth, using a 1/100 dilution of cells and the culture was incubated at 37 ^{o}C with agitation until it reached an OD_{600} of ~0.3, at which time the culture was incubated at 18^{o}C to an OD_{600} of ~0.5. At this point, 1mL was removed, centrifuged, the supernatant discarded, and the pellet was frozen for later analysis of protein expression in whole cells prior to induction. Overexpression of the protein was induced by addition of ITPG to the cell culture to a final concentration 0.1 mM. The cell culture was then left to incubate overnight (at least 8 hours) at 18 ^{o}C with agitation.

Following incubation, a sample of the overnight cell culture was then diluted with LB to an OD_{600} of approximately ~0.5, centrifuged, the supernatant discarded, and the pellet was frozen for later analysis. The remainder of the cell culture was then centrifuged at 4 ^{o}C for 10 minutes at 2,700 x *g* and the supernatant discarded. The cell pellets were resuspended in sodium-phosphate buffer (*I* = 50mM), 0.3M NaCl, pH 7.5. Approximately 6% of the original culture volume will be used for resuspension (e.g 30mL of buffer for 500mL original cell culture). Approximately 2mg each of egg white lysosyme and Bovine DNaseI were then added to the cell resuspension, and the resuspension was allowed to incubate on ice for 30 minutes. The resuspension was then sonicated on ice with 6 x 10s pulses and 50s intervening cooling periods at 40% amplitude, for a total of 60s of sonication. After the sonication step, the lysed cells were transferred into high-speed centrifuge tubes and spun at 28,000 x *g* for 25 minutes at 4 ^{o}C. The supernatant (raw extract, about 50mL) was transferred to a sterile, cold 50mL conical tube, and 100µL was removed for later analysis. A portion of the pellet was also taken for analysis of insoluble cell debris.

Exactly 2mL of Ni-NTA slurry per 500mL of original culture was added to the remaining raw extract. The raw extract was then allowed to shake gently on ice for 10 minutes. The slurry mixture was loaded onto a drip column and the flow through was collected and stored at 4 ^{o}C for further analysis. The resin was rinsed twice with 5mL of wash buffer (50mM sodium phosphate, 0.3M NaCl, 20mM imidazole, pH 7.4). Wash fractions were collected in 15mL conical tubes and stored at 4 ^{o}C for later analysis. His-tagged proteins were eluted with elution buffer (50mM sodium phosphate, 0.3M NaCl, 250mM imidazole, pH 7.4). Elution fractions were pooled into a pre-rinsed 30kDa MWCO ultrafiltration tube and concentrated by centrifugation at 4,000 x *g* for 40 minutes at 4 ^{o}C. The retained liquid was then desalted using a desalting column using 20mM Tris-Cl, pH 7.5 as equilibration buffer. The eluate from the column was then run over an anion exchange column attached to a Fast Protein Liquid Chromatography (FPLC) system. FPLC equilibration was done using 20mM Tris-Cl, pH 7.5 (buffer A) and

gradient elution was done using 0-50% 20mM Tris-Cl, 1.0M NaCl, pH 7.5 (buffer B) over three column volumes. A wash at 100% 20mM Tris-Cl 1.0M NaCl , pH 7.5 was then performed over two column volumes, followed by a re-equilibration using 20mM Tris-Cl, pH 7.5 over two column volumes. Fractions containing the desired protein were then concentrated using pre-rinsed 30k MWCO centrifugal devices to 250µL for size exclusion. Size exclusion was performed using isocratic flow at 95% buffer A, 5% buffer B over a single column volume. Fractions containing the desired protein were pooled and concentrated in a 30k MWCO centrifugal device for 40 minutes at 4,000 *x g*. Following concentration, the protein content of the sample was then quantified using the Bradford Assay.

Quantification of Protein by the Bradford Assay

Bradford Reagent was first prepared using 0.01% (w/v) Coomassie Blue G-250, 5% (v/v) ethanol, 10% (v/v) phosphoric acid diluted to an appropriate volume using ultrapure water. A standard curve was then prepared using 0, 10, 20, 30 and 40µg of bovine serum albumin (BSA) in ultrapure water. 5mL of Bradford Reagent was added to each standard sample and allowed to incubate for 6 minutes at room temperature. The samples were then analyzed spectrophotometrically at 595nm, and a standard curve was constructed through linear regression on Microsoft Excel. Using the standard curve, the protein concentration in each sample was solved. Any samples whose absorbance was outside of the dynamic range of the standard curve was diluted or re-made at a higher concentration and reanalyzed, as required.

SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gels were prepared using a TGX FastCast Stain-Free acrylamide kit using the included instructions. Samples were prepared to include approximately 4µg of protein in each well, to aid in protein purity analysis. Sterile H_2O was used to dilute samples to 16µL as required. Samples were prepared by adding 4µL of Laemmli reducing buffer (25mM Tris, pH 6.8, 20%

(w/v) glycerol, 4% (w/v) sodium dodecyl sulfate, 10mM dithiothreitol, 0.015% (w/v) bromophenol blue) to reach a total sample volume of 20µL and heated in boiling water for 5 minutes and allowed to cool at room temperature. Once cool, the samples were spun in an Eppendorf 5415C microcentrifuge at 14,000 RPM for 1 minute to remove particulates. After loading the samples into the wells, the gel was run at 100V for 25 minutes to allow the samples to stack appropriately before entering the resolving gel. The voltage was then turned up to 200V and the gel was run until the sample dye exited the bottom of the gel. The gel was then removed from the apparatus and placed onto a Bio-Rad Gel Doc EZ Stain-Free tray for gel imaging.

EH Activity Assays Towards Styrene Oxide

HPLC assays were performed as 200µL reactions containing 2mM racemic styrene oxide and 50mM sodium phosphate buffer, pH 8.0, as previously established [11]. The reaction was initiated by adding 50µM of enzyme and incubated at 25 ^{o}C for 1h. Control reactions which lack enzyme were performed in parallel. After incubation, the reaction was halted by addition of 200µL of ethyl acetate and agitated by vortexing. The reaction was then spun in a microcentrifuge at 14,000rpm for 1 minute to allow for partitioning of the organic and aqueous portion of the reaction. The organic portion was removed via micropippete and transferred to another 1.5mL microcentrifuge tube. The ethyl acetate extraction was repeated two more times, for a total of 600µL of ethyl acetate. The extracted reactions were left to air-dry overnight. The resulting residue was resuspended in 50µL of acetonitrile and analyzed in 10µL injections by HPLC with PDA detection at 254nm. Using a flow rate of 0.25mL/min, a 25% acetonitrile in water elution was for 7 minutes, followed by 3 minutes in 100% acetonitrile. The run ended with a 5 minute regeneration of the column at 25% acetonitrile in water.

Determination of EH Regioselectivity by Chiral HPLC

Similar to the EH activity assays, chiral HPLC assays was performed as 200µL reactions containing 2mM of either (*S*) or (*R*)-Styrene oxide and 50mM sodium phosphate buffer, pH 8.0. The reaction was initiated by adding 50µM of enzyme and incubated at 25°C for 1h. Control reactions which lack enzyme were performed in parallel. Additional controls were prepared by addition of (*R*) or (*S*)-1-phenyl-1,2-ethanediol to allow for analysis of product retention times. After incubation, the reaction was halted by addition of 200µL of ethyl acetate and agitated by vortexing. The reaction was then spun in a microcentrifuge at 14,000rpm for 1 minute to allow for partitioning of the organic and aqueous portion of the reaction. The organic portion was removed via micropippete and transferred to another 1.5mL microcentrifuge tube. The ethyl acetate extraction was repeated two more times, for a total of 600µL of ethyl acetate. The resulting residue was resuspended in 50µL of isopropanol and analyzed by chiral HPLC in 10µL injections with PDA detection at 254nm. Chiral HPLC analysis was performed using isocratic flow of 2.5% isopropanol in hexane for 90 minutes at 0.25mL/min. The obtained peak was then compared to the product standards to determine the extent of enantioselectivity for each EH.

Co-injection was performed by addition of 1uL of 200mM of either (*R*) or (*S*)-1-phenyl-1,2-ethanediol in isopropanol to the resuspended residue to confirm the identity of the retrieved product.

Determination of EH Enantioselectivity by Kinetic Assay

EH Kinetics were performed at 25 °C in a 1mL quartz cuvette. Reactions are performed with 10µL of 300mM $NaIO_4$ in H_2O, 25µL of 1-300mM (*R*) or (*S*)-styrene oxide in DMSO, and an appropriate amount of enzyme in 50mM $NaPO_4$, pH 8.0 to a total volume of 1mL [11]. Product diol formation by the EH is followed by the rapid oxidation of the diol to benzaldehyde by $NaIO_4$. Product formation is monitored for several minutes by the increase in absorbance at 290nm. Michaelis-Menten kinetics were fit to the data by comparing substrate concentration versus initial reaction velocity with the appropriate regression tools in R, available online at https://www.r-project.org/.

EH X-Ray Crystallography Trials

Purified native and mutant EHs were subjected to protein crystallography trials. 96-well screens were performed as sitting drops in Intelliplate 96-well plates using a crystallography robot. All crystal trays were incubated at 18 ^{o}C and checked every 72h using a microscope. Hits found in initial screens were pursued as 24-well hanging-drop screens with slight variations to the initial hit. Diffraction of protein crystals was initially monitored at the University of Waterloo (Waterloo, ON), and crystals found to diffract sufficiently were sent to the Canadian Light Source (CLS, Saskatoon, SK). Molecular replacement and iterative rounds of structure refinement were performed using Phenix[34] and Coot[35].

Results

Bioinformatics analysis

Alignment of the EH sequences from the cluster of tryptophan-containing EHs reveals several amino acid sites which are conserved throughout the cluster (**Figure 5**). First, it is of note that the predicted protein M0B613 from *Natrialba asiatica* is significantly divergent from the other tryptophan-containing EHs, with large gaps appearing in the Clustal Omega-guided sequence alignment, despite showing the catalytic triad and epoxide binding residues at the expected positions. Furthermore, the predicted protein from *N. asiatica* is significantly shorter than the other tryptophan-containing EHs, which is unsurprising, as N. asiatica is an archaeal protein. These characteristics of this predicted protein make it difficult to compare to the other enzymes found in this cluster. Close analysis of the remaining EHs within this cluster reveals several interesting residues which are only conserved in tryptophan-conatining EHs. When mapped onto the solved crystal structure of SghF, several amino acid sites such as A194, Q303, and Q370 (SghF numbering) appear to be distal from the active site of EHs (**Figure 5**). Of these residues, Q303 merits further investigation, as this position is also conserved as either a histidine

or an arginine in tryptophan-containing EHs. The mutation from a basic amino acid to a polar uncharged amino acid may affect the cataytic cycle. Interestingly, a strongly conserved stretch of residues L187, A188, D189, P190, E191, and H192 (**Figure** 5) present in tryptophan-containing EHs distal from the active sites residues as well. Despite their distance from the active site, these residues are compelling candidates for future engineering efforts in order to fine tune these EHs, as it has been found in many cases that amino acids that are distal from the active site can affect catalysis. The residues Y138, F229, and L304 are also conserved and are found proximal to the active site. Y138 is found quite close to the epoxide-binding tyrosine-tryptophan pair, and is mostly conserved as a phenylalanine in tyrosine-containing EHs. The phenylalanine in postion 229 is found beside the epoxide binding tryptophan, and is mostly variable in tyrosine containing EHs. Finally, L304 is found immediately beside the invariable epoxide-binding tyrosine, and once again is mostly variable in predicted retaining EHs. Altogether, the conserved residues found within the sequence alignment merit further investigation, as these amino acids which are conserved in the inverting EHs may hold the key to uncovering the molecular determinants of the regioselectivity of epoxide hydrolases.

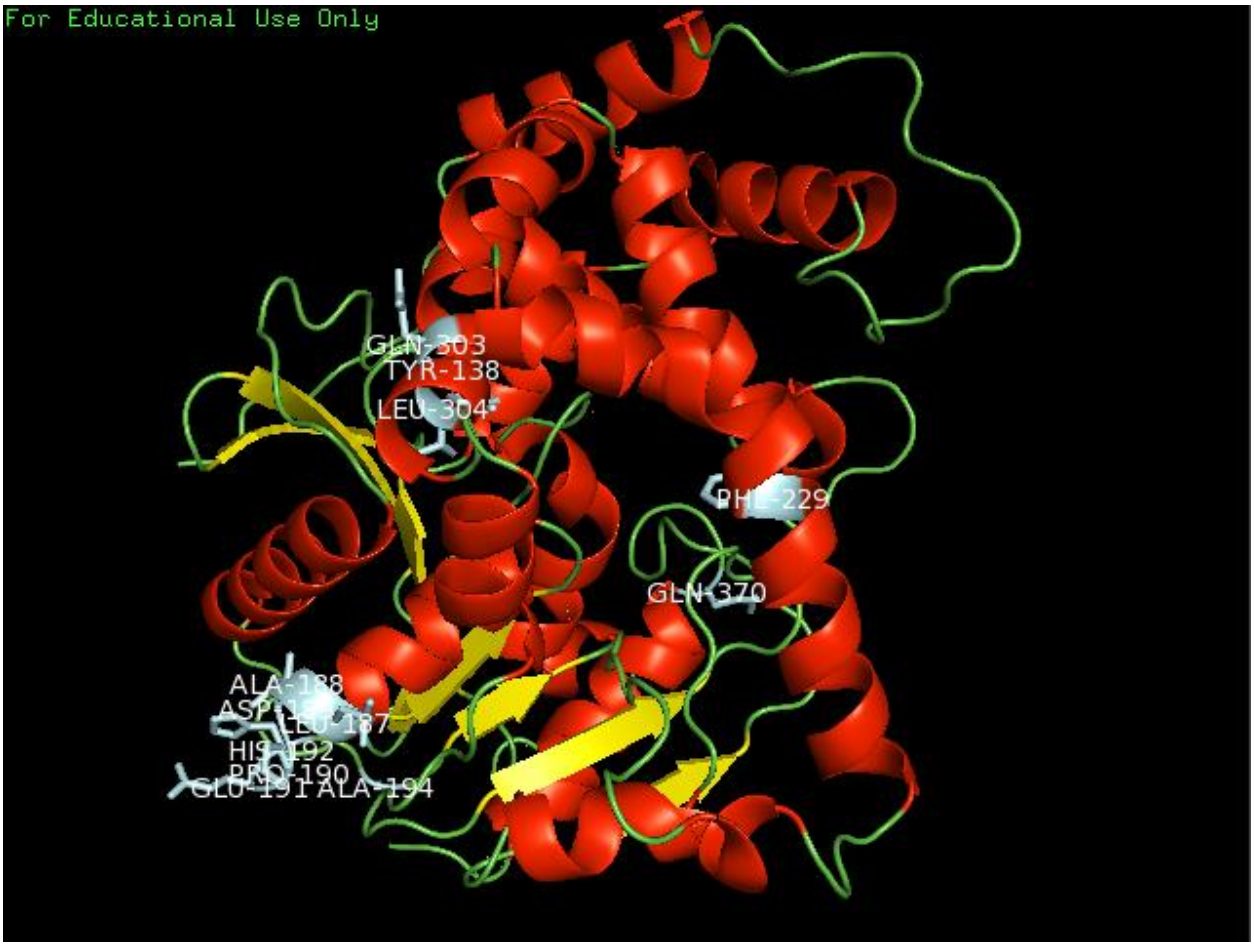


Figure 5. Cartoon representation of the conserved residues within the tryptophan-containing EHs shown on SghF. Conserved residues are shown in white.

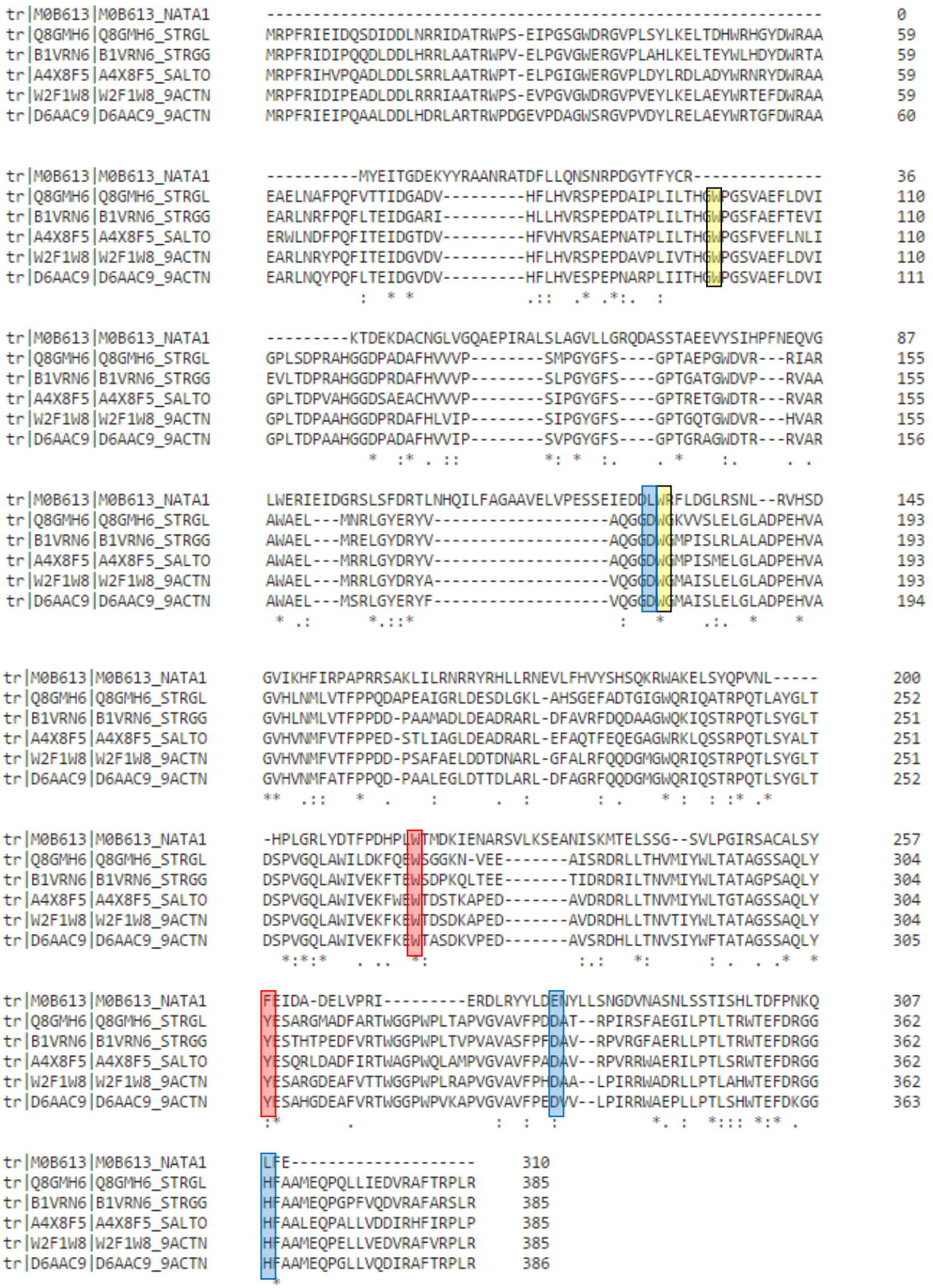
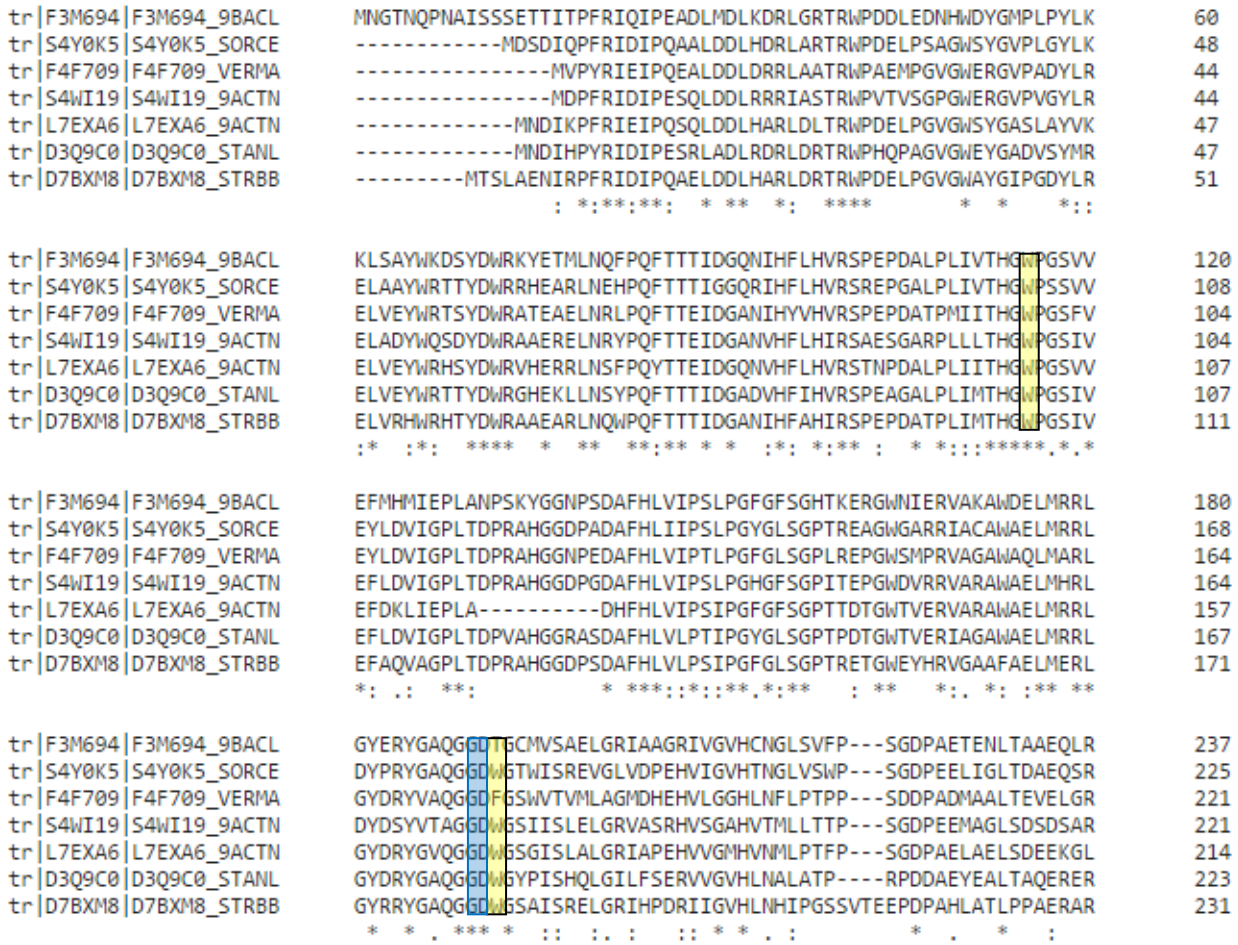

tr|M0B613|M0B613_NATA1
tr|Q8GMH6|Q8GMH6_STRGL
tr|B1VRN6|B1VRN6_STRGG
tr|A4X8F5|A4X8F5_SALTO
tr|W2F1W8|W2F1W8_9ACTN
tr|D6AAC9|D6AAC9_9ACTN

Figure 6. Alignment of tryptophan-containing EHs by Clustal Omega. Conserved residues are indicated with asterisks. Oxyanion hole residues W100 and W177 are highlighted in yellow. Catalytic triad residues D176, D337, and H364 are highlighted in red. Epoxide-binding residues W236 and Y304 are highlighted in blue (SghF numbering).

Similarly, alignment of the EH sequence that surround the tryptophan-containing EHs reveal several other amino acid sites which are conserved throughout (**Figure 6**). First, the Acid/His/Acid catalytic triad found in all α/β-fold epoxide hydrolases was confirmed to be present, at D176, H364, and D336 (SghF numbering). Finally, the epoxide-binding residues were found at W236, and Y305. Similarly, the tyrosine-containing EHs found proximal to the tryptophan-containing cluster were aligned, and the Acid/His/Acid catalytic triad was found at D175, H363, and D335, and the epoxide-binding residues were found at Y235 and Y304 (CynF numbering).



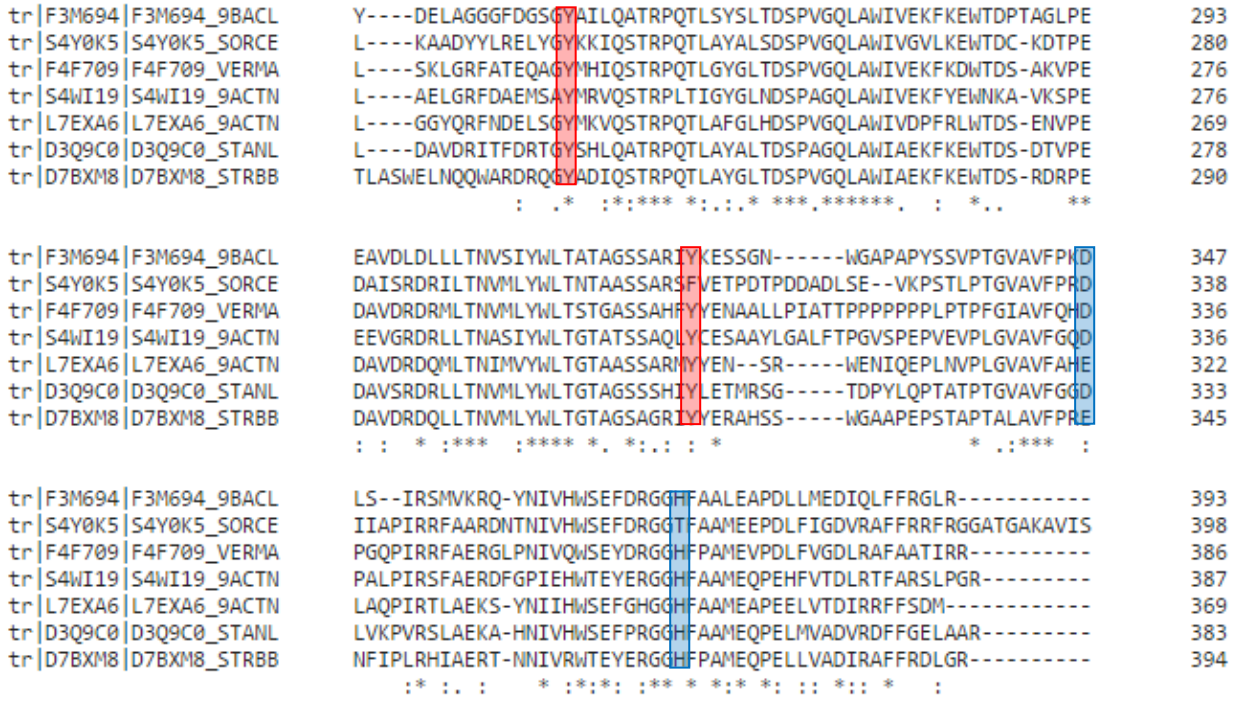


Figure 7. Alignment of tyrosine-containing EHs by Clustal Omega. Conserved residues are indicated by asterisks. Oxyanion hole residues W99 and W176 are highlighted in yellow. Catalytic triad residues D175, D336, and H363 are highlighted in red. Epoxide-binding residues W235 and Y303 are highlighted in blue (CynF numbering).

Amplification and Cloning of the Genes that Encode SghF and CynF

The CynF gene was amplified as described using the primers CynF-NdeI-F and CynF-XhoI-R from cosmid GD12, obtained from Amy Lane, University of North Florida. PCR products were run on a 1% agarose gel. Amplicons of the expected size of ~1.2kb corresponding to the gene encoding CynF were observed throughout the annealing gradient (**Figure 7**). The highlighted amplicon was extracted from the gel and used for ligation into pGEM-T-Easy after addition of the adenine overhang.

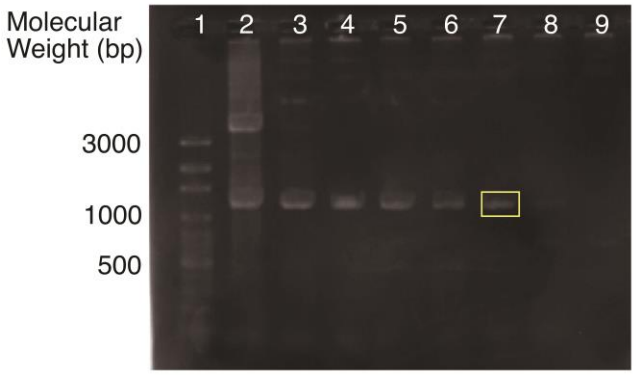


Figure 8. Image of PCR amplification of the genes encoding for cynF run on a 1% (w/v) agarose gel. Annealing gradients of 50-65 ^{o}C were used and are loaded in ascending order left to right. 2µL of Axygen 100bp DNA ladder is loaded in lane 1. The highlighted amplicon was extracted and used for ligation into pGEM-T Easy.

The resultant DNA prep was digested with *EcoRI* and run on a 1% (w/v) agarose gel (**Figure 8**) to verify the size of the insert. Plasmid DNA corresponding to the highlighted insert was confirmed by DNA sequencing, yielding the construct cynF-pGEM-T Easy, with an encoded 5' NdeI restriction site upstream of the N-terminal MAH_6VD_4K fusion tag used in previous EH studies [13] and a 3' XhoI restriction site to allow for easy cloning into pET29.

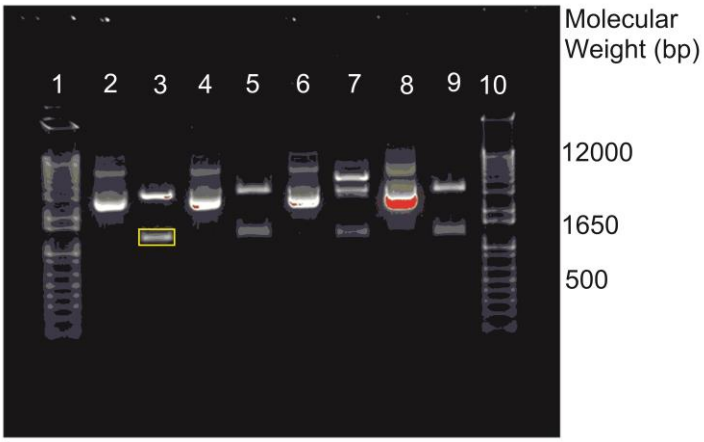


Figure 9. Image of EcoRI digests of cynF-pGEM-T Easy run on a 1% (w/v) agarose gel. 2µL of Invitrogen 1kb+ ladder is loaded in lanes 1 and 10. Plasmid DNA corresponding to the highlighted inserts were sent to The Center for Applied Genomics for sequence confirmation.

The confirmed CynF-pGEM-T Easy construct was then subcloned into pET29, yielding the expression construct CynF-pET29, encoding the CynF gene with an N-terminal MAH_6VD_4K fusion tag to allow for both Ni-NTA affinity purification as well as cleavage of the tag by enterokinase.

The SghF gene was received as a synthesized pUC57 construct from BioBasic (Markham, ON) with restriction enzyme sites for NdeI and XhoI at the 5' and 3' end of the gene respectively. Both the SghF-pUC57 construct as well as an empty pET28 construct were digested with NdeI and XhoI to produce the necessary sticky ends to allow for ligation (**Figure 9**). This yielded the expression construct SghF-pET28, encoding the SghF gene with an N-terminal His_6 tag upstream of a thrombin cleavage site, allowing for Ni-NTA affinity purification as well as cleavage of the tag by thrombin.

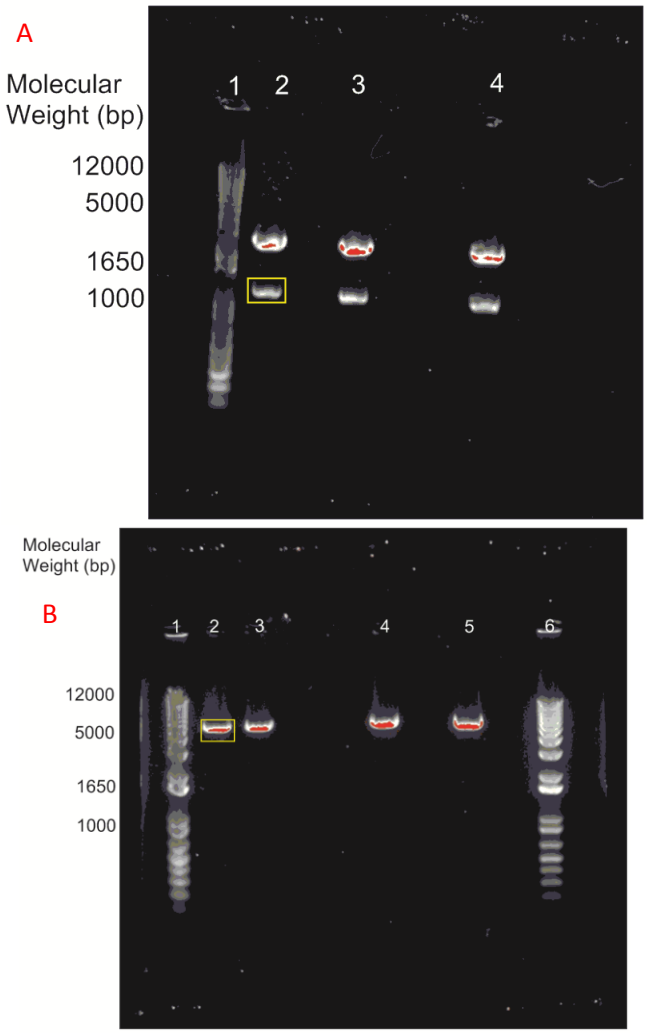


Figure 10. Agarose gel image of (A) SghF-pUC57 digested with NdeI and XhoI and (B) pET28 vector digested with NdeI and XhoI. Highlighted bands were excised from the gel for ligation.

Generation of the EH mutants mutants SghF-H364Q, SghF-H364A, SghF-D176N, CynF-H363Q, CynF-H363A, and CynF-D176N through overlap extension PCR

Primers were designed based on the catalytic residues of CynF and SghF in order to perturb the native mechanism of the EHs. Mutants for CynF were prepared as previously described using overlap extension PCR. The expression construct CynF-pET29 as the template for the initial round of PCR, along with the appropriate pair of primers to yield the desired mutant. Once both the upstream and downstream fragments of the gene were prepared, a second round of PCR was performed using the 5' and 3' primers. The mature PCR product was then run on a 1% (w/v) agarose gel to verify the size of the amplicon. Once verified, the amplicon was excised from the gel and ligated into pET29 for protein expression. Mutants for SghF were prepared in an identical manner; SghF-pET28 was used as the template for the initial round of PCR (**Figure 10 (A)**), and the final verified amplicon was excised from the gel and ligated into pET28 for protein expression (**Figure 10 (B)**). The constructs SghF-H364A-pET28, SghF-H364Q-pET28, SghF-D176N-pET28, CynF-H363A-pET29, CynF-H363Q-pET29, and CynF-D175N-pET29 were sent for DNA sequencing to confirm both the identity of the insert and the incorporation of the mutation.

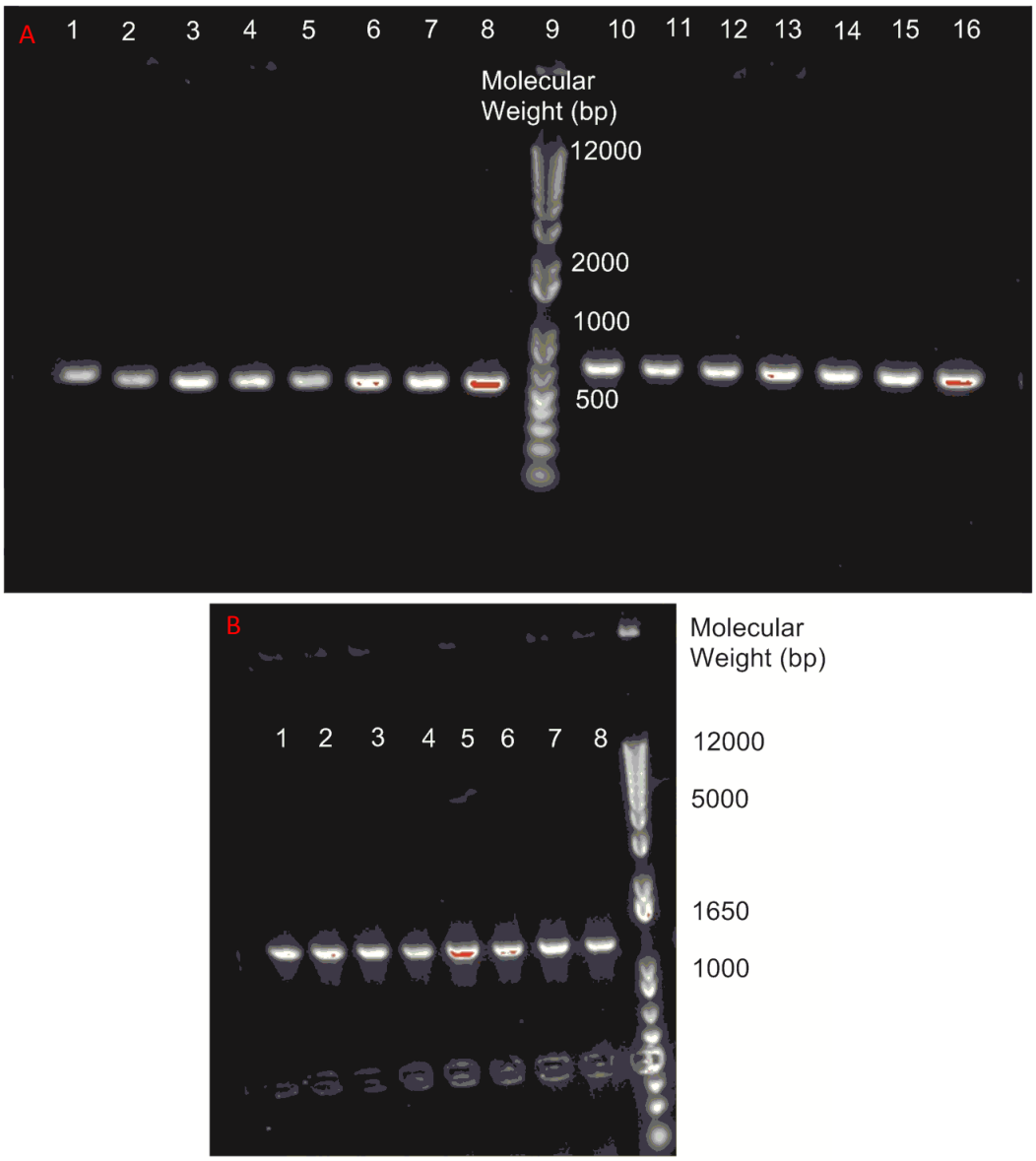


Figure 11. **(A)** PCR amplification of SghF-D176N. The portions of the gene upstream (lanes 1-8) and downstream (lanes 10-16) of the mutation are shown. Homologous amplicons were pooled together for excision from the gel to improve DNA yield. 2µL of Invitrogen 1kb+ ladder were used in lane 9 **(B)** Full extension amplification of the mutant SghF-D176N (lanes 1-8). Homologous amplicons were pooled together for excision from the gel to improve DNA yield. 2µL of Invitrogen 1kb+ ladder were used in lane 9.

Expression and Purification of the EHs SghF and CynF, and the mutants SghF-H364Q, SghF-H364A, SghF-D176N, CynF-H363Q, CynF-H363A, and CynF-D176N

After transformation of the appropriate expression construct into *E.coli* BL21, SghF (13.5mg/L), SghF-H364Q (2.5mg/L), SghF-H364A (0.23mg/L), SghF-D176N (11.8mg/L), CynF (20.3mg/L), CynF-H363Q

(3.2mg/L), CynF-H363A (0.1mg/L) and CynF-D175N (0.6mg/L) were purified to near homogeneity as N-terminal His_6-tagged proteins by Ni-NTA affinity column chromatography, anion exchange chromatography, and gel filtration chromatography as necessary. SDS-PAGE showed a single protein band consistent with the expected molecular weight of all of the EHs purified (~45kDa).

HPLC Activity Assay for SghF and CynF and their Mutants

Due to unavailability of the natural substrates for CynF and SghF, styrene oxide was used as a substrate analogue to test their activity. A single peak with a retention time identical to that of authentic 1-phenyl-1,2-ethanediol was observed in both cases (**Figure 11** and **Figure A6** in the Appendix).

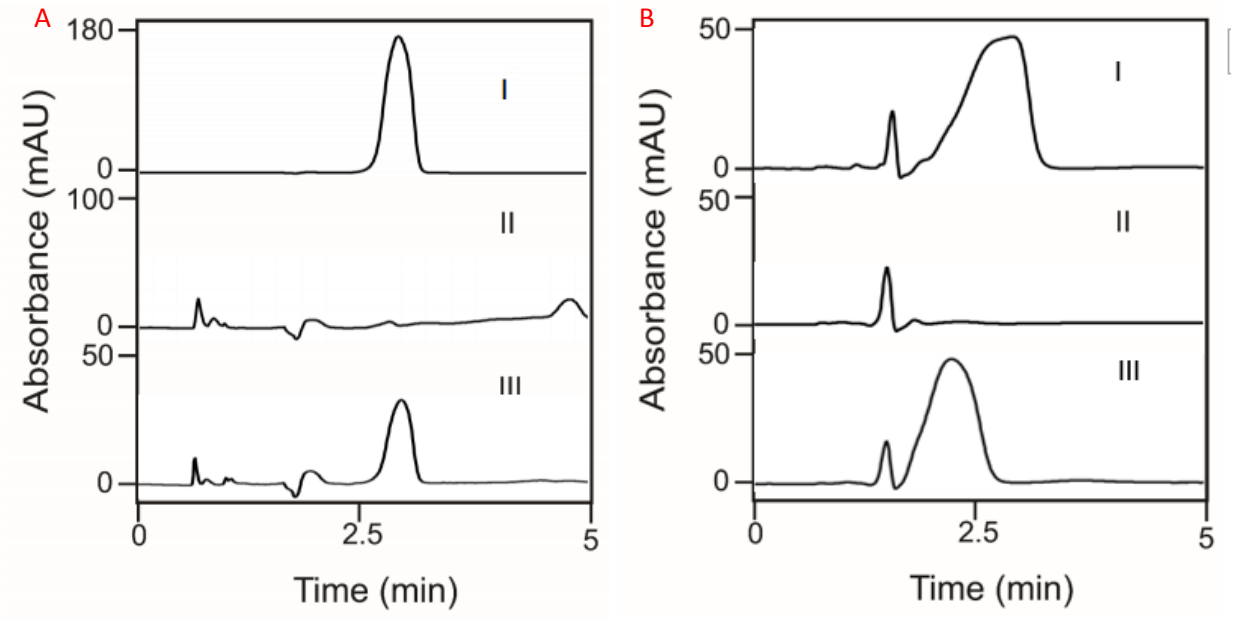


Figure 12. HPLC activity assay for hydrolysis of racemic styrene oxide by (A) CynF and (B) SghF. (I) 1-phenyl-1,2-ethanediol standard (II) control with no enzyme (III) styrene oxide treated with 50µM CynF

The reverse-phase HPLC assay performed was done in order to determine whether the purified EHs were active towards styrene oxide. Close analysis of **Figure 11** and **Figure A6** in the Appendix reveals that the HPLC profile for styrene oxide treated with either SghF or CynF provides a similar retention time to that observed for authentic 1-phenyl-1,2-ethanediol. Although it is expected to exhibit absorbance at

254nm and therefore provide an observable peak, there was not a significant peak that can be attributed to styrene oxide in any of the control reactions. The absence of a peak can be attributed to the evaporation of styrene oxide under low pressure conditions.

Enantioselectivity of CynF and SghF

Enantiomerically pure (*R*)- and (*S*)- styrene oxide were used to analyze the enantioselectivity of *CynF* and *SghF* using chiral HPLC. Authentic (*R*)- and (*S*)- 1-phenyl-1,2-ethanediol were used to determine the retention times for each product enantiomer (**Figure 12**). As the retention times for the authentic standards differed from the eznyme-prepared products, co-injections were performed as described in the Methods section. As expected, it was found that hydrolysis of (*S*)-styrene oxide by CynF resulted in the retention of the stereocenter in the diol product, yielding (*S*)-1-phenyl-1,2-ethanediol (pane IV in **Figure 12**). In addition, it was determined that hydrolysis of (*S*)-styrene oxide by SghF resulted in the inversion of the stereocenter in the diol product, yielding (*R*)-1-phenyl-1,2-ethanediol (pane I in **Figure 12**). The retention times of the authentic (*R*)- or (*S*)-1-phenyl-1,2-ethanediol were significantly different when compared to those for the enzyme hydrolysis product run on the same Chiralcel OD-H column. As such, it was deemed necessary to perform a co-injection of the enzyme-catalyzed preparations with the expected authentic diol. After co-injection, it was found that CynF-catalyzed hydrolysis of (*S*)-styrene oxide yielded (*S*)-1-phenyl-1,2-ethanediol (pane V in **Figure 12**), and SghF-catalyzed hydrolysis of (*S*)-styrene oxide yielded (*R*)-1-phenyl-1,2-ethanediol. CynF-catalyzed hydrolysis of (*R*)-styrene oxide provided a mix of approximately two parts (*R*)- and one part (*S*)-1-phenyl-1,2-ethanediol, which is consistent with results found for other EHs when hydrolysing (*R*)-styrene oxide [11]. Interestingly, SghF-catalyzed hydrolysis of (*R*)-styrene oxide yielded little to no product diol, suggesting that the kinetics for SghF-catalyzed hydrolysis of styrene oxide may strongly favour the (*S*)-enantiomer.

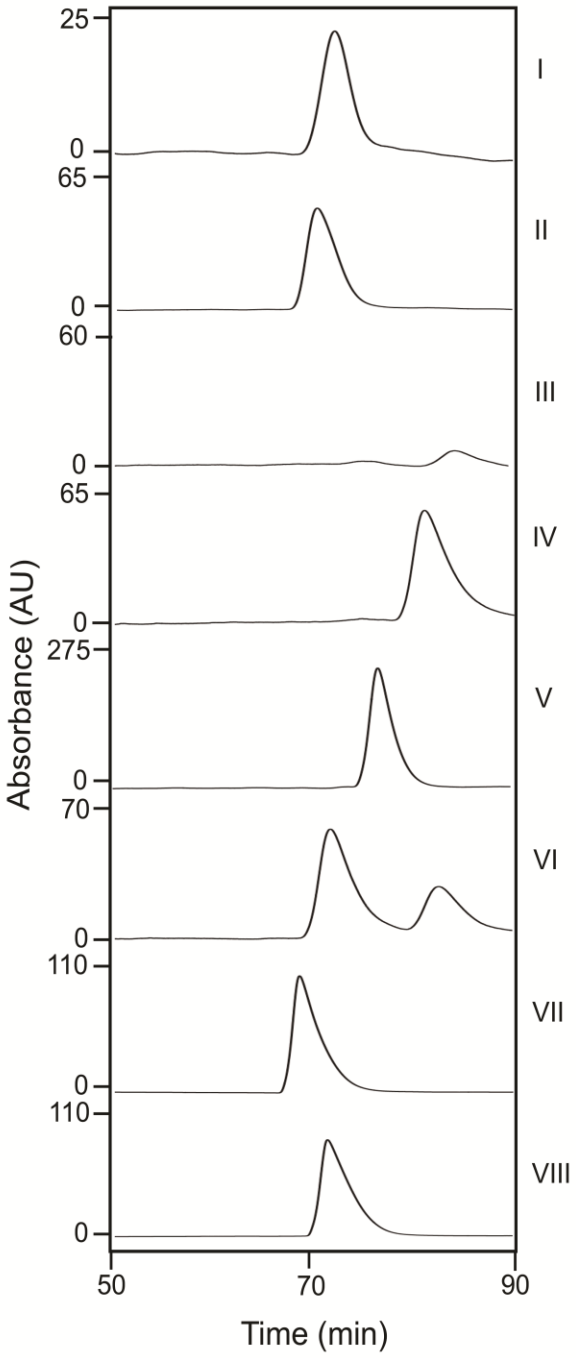


Figure 13. HPLC assay for determining the stereochemistry of 1-phenyl-1,2-ethanediol as a result of catalysis by CynF or SghF (I) SghF-catalyzed hydrolysis of (S)-styrene oxide, (II) SghF-catalyzed hydrolysis of (S)-styrene oxide co-injected with authentic (R)-1-phenyl-1,2-ethanediol (III) SghF-catalyzed hydrolysis of (R)-styrene oxide, (IV) CynF-catalyzed hydrolysis of (S)-styrene oxide, (V) CynF-catalyzed hydrolysis of (S)-styrene oxide co-injected with authentic (S)-1-phenyl-1,2-ethanediol, (VI) CynF-catalyzed hydrolysis of (R)-styrene oxide, (VII) (R)-1-phenyl-1,2-ethanediol standard, (VIII) (S)-1-phenyl-1,2-ethanediol standard.

Kinetics of CynF and SghF and their mutants towards (*R*)- and (*S*)- styrene oxide

Michaelis-Menten kinetics parameters were determined (**Table 4**) for the native enzymes *CynF* and *SghF* using a previously established assay [11,36]. The activity of the mutants SghF-H364Q, SghF-H364A, SghF-D176N, CynF-H363A, CynF-H363Q, and CynF-D175N were also analyzed through this assay.

Table 4. Steady-state kinetic parameters for CynF and SghF and variants and previously characterized endiyne-associated EHs towards (R)- and (S)- styrene Oxide as substrates.

	(*R*)-**Styrene Oxide substrate**			(*S*)-**Styrene Oxide substrate**			
EH	K_{cat}(min^{-1})	K_m (mM)	K_{cat}/K_m	K_{cat}(min^{-1})	K_m (mM)	K_{cat}/K_m	*E*
SghF wild-type	ND	ND	0.29±0.045	70.3 ± 1.7 min^{-1}	0.37 ± 0.036	190	655
SghF-H364A	ND	ND	ND	ND	ND	ND	-
SghF-H364Q	ND	ND	ND	ND	ND	ND	-
SghF-D176N	ND	ND	0.074±0.014	ND	ND	0.14±0.056	1.89
CynF wild-type	ND	ND	0.11±0.013	ND	ND	0.61±0.16	5.55
CynF-H363A	ND	ND	ND	ND	ND	ND	-
CynF-H363Q	ND	ND	ND	ND	ND	ND	-
CynF-D176N	ND	ND	ND	ND	ND	ND	-
SgcF[a]	7.5±0.7	2.7±1.1	2.68	43±4	0.88±0.45	53.9	20.1
SpoF[a]	3.4±0.4	2.8±0.9	1.25	34±3	1.1±0.6	32.3	25.8
SgrF[a]	7.9±0.6	2.6 ±0.6	3.11	43±4	0.88±0.45	48.9	15.7
NcsF2[b]	133±4	0.5±0.1	266	31±2	5.0±0.6	6.0	-44
KedF[c]	35±2.4	3.5±0.64	10.0	36.6±1.1	0.91±0.1	40.2	4.0

ND= Not determined

E refers to the enantioselectivity [K_{cat}/K_m(fast)]/[K_{cat}/K_m(slow)] where for positive values (*S*)-styrene oxide is the preferred (fast) substrate and for negative values (*R*)-styrene oxide is the preferred substrate.

[a]Values obtained from reference. [13]

[b]Values obtained from reference. [12]

[c]Values obtained from reference. [37]

Interestingly, kinetic characterization of CynF reveals that is able to hydrolyze both the (*S*)- and (*R*)- epoxide, with an apparent K_{cat}/K_m of 0.61±0.16 and 0.11±0.013, respectively. CynF has an increased catalytic efficiency for the (*S*)-styrene oxide (E=5.55, in favour of the (*S*)-epoxide), which agrees with the hypothesis that the (*S*)-epoxide is the natural and therefore the favoured substrate for enediyne-associated epoxide hydrolases. Kinetic characterization of SghF reveals that it efficiently hydrolyzes the substrate analogue (*S*)-styrene oxide, with an apparent K_m of 0.37 ± 0.036 mM and K_{cat} of 70.3 ± 1.7 min^{-1}, yielding a K_{cat}/K_m of 190. SghF therefore represents the most efficient of the now 7 characterized enediyne-related epoxide hydrolase for (*S*)-styrene oxide hydrolysis[13]. Interestingly, SghF can also hydrolyze (*R*)-styrene oxide, with an apparent Kcat/Km of 0.29±0.045 (**Table 4**). This reduced catalytic efficiency for the (*R*)-epoxide once again coincides with the hypothesis that the (*S*)-epoxide is the natural substrate for enediyne-associcated epoxide hydrolases [11]. SghF strongly prefers hydrolysis of the (*S*)-epoxide, with an E-value of 655 in favour of the (*S*)-epoxide.

Structural Characterization of SghF and CynF and their Mutants

SghF protein crystals were obtained by vapor diffusion against 500µL of reservoir solution in a 3µL hanging drop at 18^{o}C. The SghF protein was prepared in 50mM Tris buffer at pH 8.0 for initial crystal screening, and drops were prepared by mixing the reservoir solution with SghF in either a 2:1 or 1:2 ratio. After harvesting, crystals were soaked in one of two cryprotectant solutions: 25% (w/v) glycerol, 20% (w/v) poly(ethylene) glycol 3.35k, 0.1M Bis-Tris, pH 6.5, or 27.5% (w/v) glycerol, 20% (w/v) poly(ethylene) glycol 3.35k, 0.1M Bis-Tris, pH 6.5. The crystals were then flash frozen by rapid submersion into a liquid nitrogen bath. Diffraction data were collected at the Canadian Light source and refined to 2.7Å using the Coot and Phenix software packages (**Figure 12**).

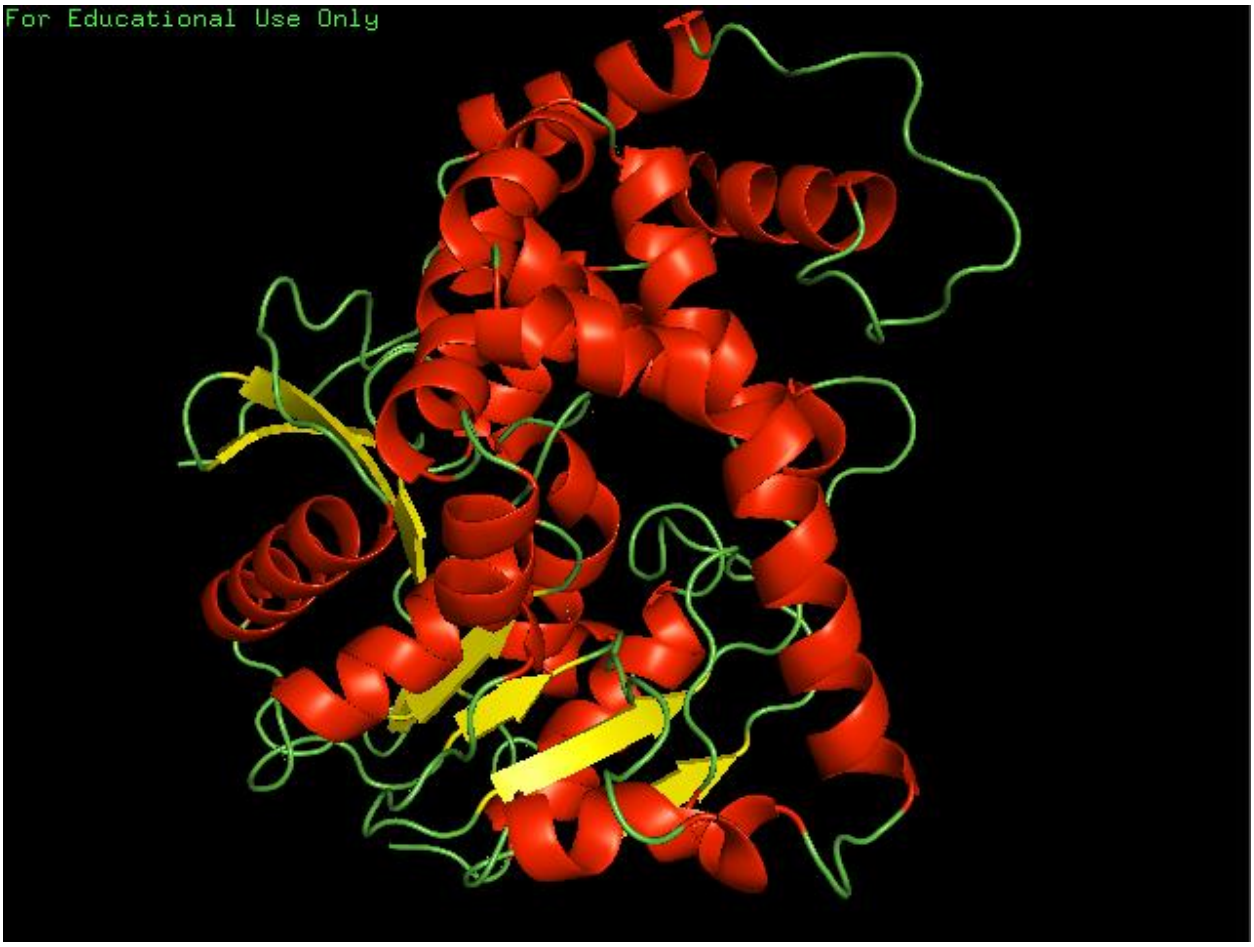


Figure 14. Cartoon representation of the solved crystal structure for SghF.

CynF and the mutants CynF-H363A, CynF-H363Q, and CynF-D175N were all subjected to 96-well crystal screening as described in the Methodology section. Despite a large number of crystal screens employed, there were no hits to follow up for any of CynF or its mutants.

Discussion

Two closely related epoxide hydrolases, CynF and SghF, were selected for bioinformatics analysis and biochemical characterization. It was expected that characterization of these enzymes may allow for the determination of the molecular determinants of epoxide hydrolase regio- and enantioselectivity. Although these enzymes are highly related, they were expected to exhibit opposing

regioselectivities as expected from the predictive model for enediyne epoxide hydrolase regio- and enantioselectivity, first proposed by Horsman et al. [13].

Once the catalytic triad residues were determined for both CynF and SghF, it was decided that several mutants were to be made. Previous work had suggested that the tyrosine to tryptophan mutation found in inverting enediyne EHs was not involved in the retention or inversion of the stereocenter in the diol product [13]. As such, it was thought that generation would provide a platform which would allow for the probing of the epoxide hydrolase catalytic mechanism.

As previously discussed, the catalytic mechanism proceeds in what is essentially two steps: formation of the akyll-enzyme complex by nucelophilic attack at one of two positions in the epoxide, and activation of water by a His-Acid pair and subsequent hydrolysis of the diol product. By mutating the residues that are thought to be involved in EH catalysis, theoretically it would be possible to probe the catalytic mechanism of EH catalysis and allow for further mechanistic and modelling studies. It has been found in the past that mammalian soluble EHs accumulate up to >60% of the alkyl-enzyme intermediate when subjected to crystal soak-ins [38,39]. While the rate of nucleophilic attack is significantly higher than the rate of hydrolysis for mammalian EHs, in some cases the level of intermediate in bacterial EHs can accumulate as little as <1% [40]. As such, mutations which selectively impede or impair the hydrolysis step of EH catalysis should greatly increase the accumulation of the alkyl-enzyme intermediate.

The first site that was selected for mutation was the catalytic triad histidine (H364 in SghF and H363 in CynF). By mutating this histidine, it is proposed that the enzyme would be unable to hydrolyze the diol product, leading to an accumulation in the alkyl-enzyme adduct (**Figure 13 (A)** and **Figure A13 (A)** in the appendix) . By mutating this amino acid to glutamine, it was thought that not only would the relatively isosteric mutation be tolerated structurally, but also the loss of the imidazole ring would disable the enzyme's ability to relay charge and activate the water molecule that is required for

hydrolysis. An alanine mutant was also made at this position, due to its non-bulky, chemically inert methyl functional group. It was also expected that this mutant would be unable to activate the water molecule that is required for hydrolysis (**Figure 13 (B)**).

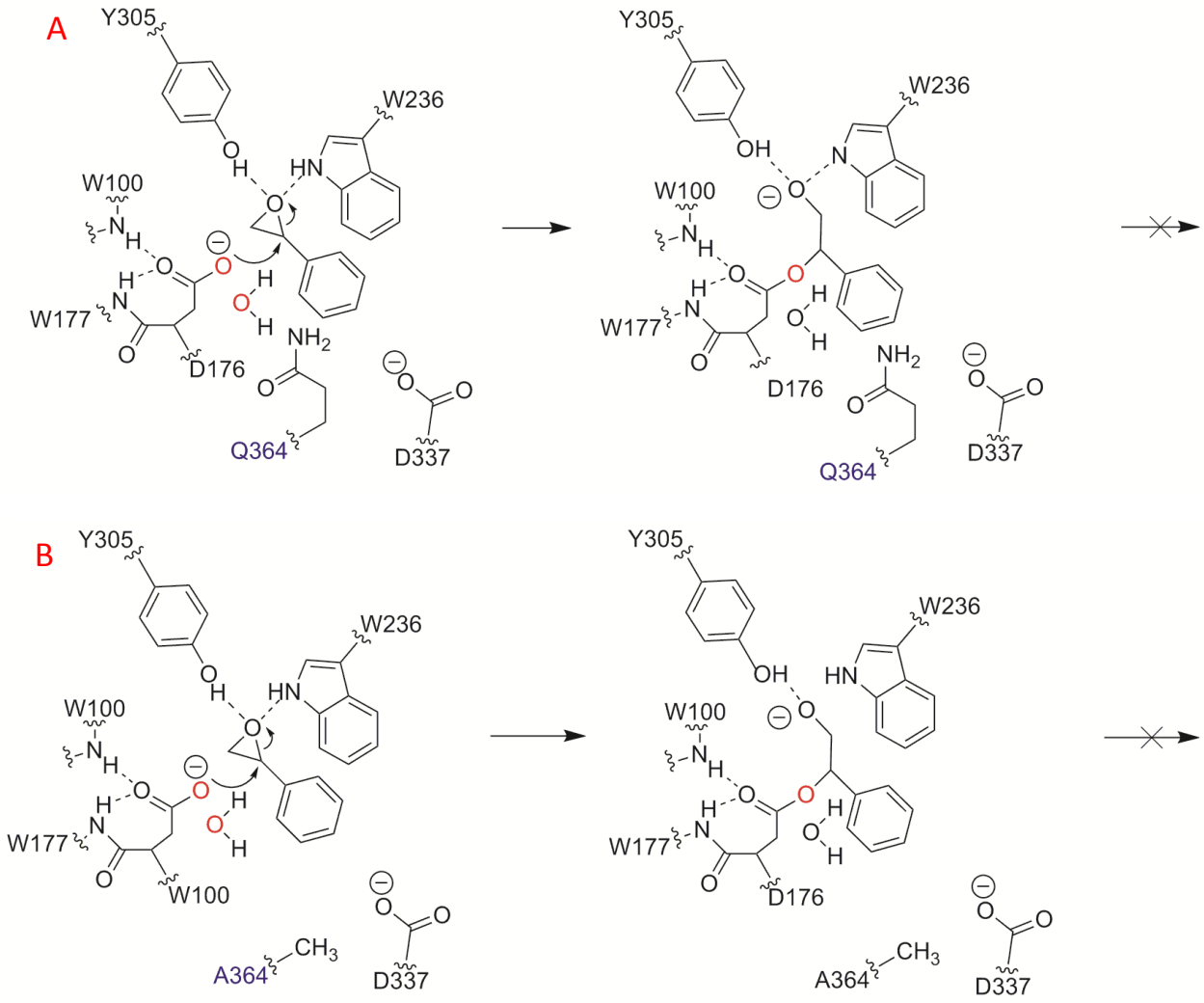


Figure 15. Proposed mechanism for styrene oxide hydrolysis by (A) SghF-H364Q and (B) SghF-H364A.

In addition, the aspartate responsible for the initial nucelophilic attack in the first step of EH hydrolysis was targeted for mutagenesis. By mutating this residue, the mutants should lose their ability to perform a nucleophilic attack on the epoxide carbon, which may lead to an accumulation of the substrate-bound form of the enzyme (**Figure 14** and **Figure A15** in the appendix). The isosteric muation

of this aspartate to asparagine should yield an epoxide hydrolase with a charge at that positon that opposes the original charge of the aspartate residue.

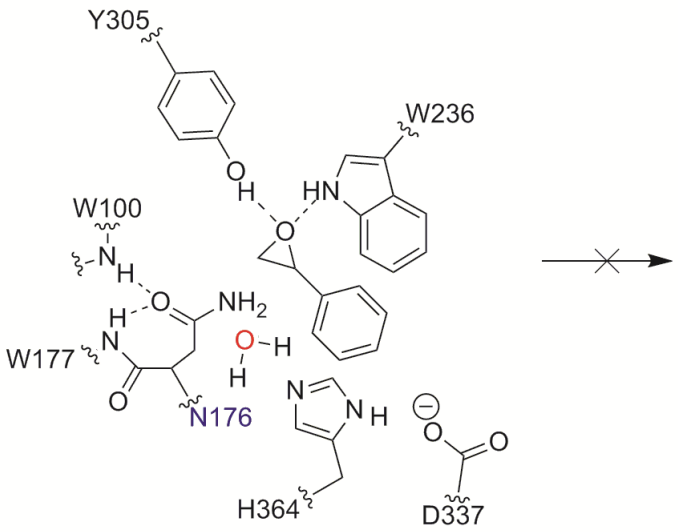


Figure 16. Proposed mechanism of styrene oxide binding by SghF-D176N

Interestingly, although no activity was found for the mutants CynF-H363A, CynF-H363Q, CynF-D176N, SghF-H364A, or SghF-H364Q, SghF-D176N was found to have some activity towards both enantiomers of styrene oxide (**Table 4**). There are currently three hypotheses which may explain this unexpected activity. The first possibility is that a lone pair of electrons from the amine nitrogen allows for the activation of the neighbouring oxygen, resulting in a nucleophilic attack and an alkyl-enzyme intermediate closely resembling the intermediate in the native enzyme's natural catalytic cycle (**Figure 15**). The enzyme is then able to hydrolyze the alkyl enzyme, similar to the natural catalytic cycle, resulting in a product diol.

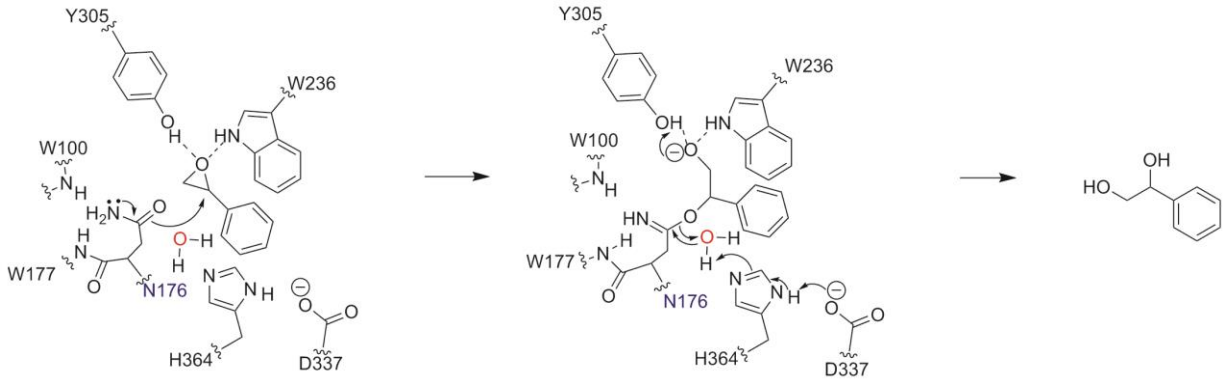


Figure 17. First hypothesis for observed hydrolase activity in SghF-D176N.

The second hypothesis for the observed SghF-D176N activity is the direct nucleophilic attack by a lone pair from the amine nitrogen of the aspargine residue. Once again, an alkyl-enzyme intermediate that closely resembles the native intermediate is formed, and hydrolyzed, resulting in a product amine. Interestingly, while sodium periodate selectively oxidizes vicinal diols, it has been shown that it is capable of oxidizing amines to some extent [41]. While the identity of the oxidized amine product is unknown, it still may absorb somewhat at 290nm, resulting in the detection of apparent activity for SghF-D176N.

While a viable hypothesis, this mechanism can be eliminated. If this mechanism were correct, then after the first catalytic cycle, the asparagine residue would be reverted back to its native aspartate (**Figure 16**). Accompanying this would be an observed increase in the catalytic rate that would be comparable to the native SghF. However, no increase in apparent activity could be observed for SghF-D176N over the time frame of the kinetics reaction. This suggests that this mechanism is not indicative of the apparent activity of SghF-D176N.

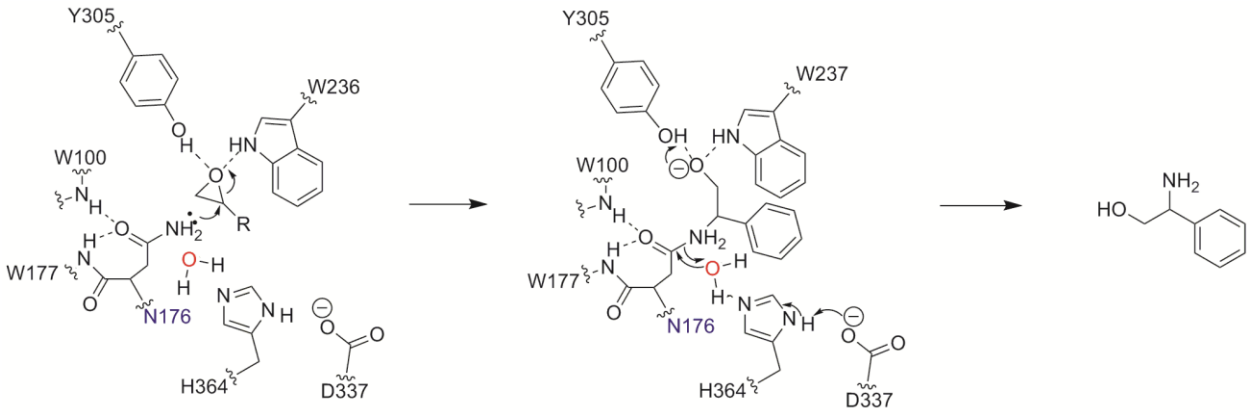


Figure 18. Second hypothesis for observed hydrolase activity in SghF-D176N

The final hypothesis involves the direct activation of water by H364. In one step, the product diol is formed through a general base mechanism (**Figure 17**).

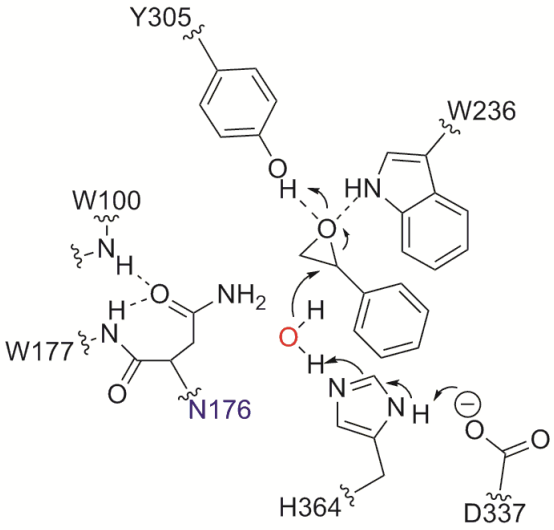


Figure 19. Third hypothesis for observed activity in SghF-D176N

Interestingly, kinetic characterization of CynF reveals that is able to hydrolyze both the (*S*)- and (*R*)- epoxide, with an apparent K_{cat}/K_m of 0.61±0.16 and 0.11±0.013, respectively. CynF has an increased catalytic efficiency for the (*S*)-styrene oxide (E=5.55, in favour of the (*S*)-epoxide), which agrees with the hypothesis that the (*S*)-epoxide is the natural and therefore the favoured substrate for enediyne-associated epoxide hydrolases. Kinetic characterization of SghF reveals that it efficiently hydrolyzes the

substrate analogue (*S*)-styrene oxide, with an apparent K_m of 0.37 ± 0.036 mM and K_{cat} of 70.3 ± 1.7 min^{-1}, yielding a K_{cat}/K_m of 190. SghF therefore represents the most efficient of the now 7 characterized enediyne-related epoxide hydrolase for (*S*)-styrene oxide hydrolysis[13]. Interestingly, SghF can also hydrolyze (*R*)-styrene oxide, with an apparent Kcat/Km of 0.29±0.045 (**Table 4**). This reduced catalytic efficiency for the (*R*)-epoxide once again coincides with the hypothesis that the (*S*)-epoxide is the natural substrate for enediyne-associcated epoxide hydrolases [11]. SghF strongly prefers hydrolysis of the (*S*)-epoxide, with an E-value of 655 in favour of the (*S*)-epoxide. This E-value is the largest by one order of magnitude for any of the characterized enediyne-associated epoxide hydrolases for either enantiomer of styrene oxide. Thus, SghF may serve as a sensible starting point for future mechanistic and biochemical studies in order to produce an efficient biocatalyst for the regioselective production of 1-phenyl-1,2-ethanediol from styrene oxide.

The regio- and enantioselectivity of EHs provide an intriguing platform for many high-demand industries that employ traditional chemical synthesis. For instance, EHs are potentially useful in the pharmaceutical industry, where many enzymes are already employed as catalysts in the production of drugs whose value in sales often exceed millions of dollars worldwide annually. For example, semisynthetic penicillins and substrates for the synthesis of other β-lactams can be synthesized using the enzyme penicillin acylase[42]. As a whole, the pharmaceutical industry is experiencing a trend towards the use of biopharmaceuticals. In 2010, five of the top ten grossing drugs worldwide were drugs that were produced from biological sources [43]. In addition, EHs may also be useful in the medical industry as a diagnostic test, taking advantage of their regioselectivity and ability to hydrolyze a wide variety of substrates efficiently. Enzymes are already widely employed in diagnostic testing, the most well-known being the use of glucose oxidase in urine testing [44].

The work described herein sets the table for future studies involving the probing of the α/β-epoxide hydrolase mechanism for styrene oxide. Kinetic and regioselectivity characterization of both

CynF and SghF has determined that SghF may be a robust starting platform for engineering efforts to generate enantiomerically pure diols through biocatalysis. SghF represents the most efficient (K_{cat}/K_m of 190 for (*S*)-styrene oxide) and enantioselective (E=655, in favour of the (*S*)-epoxide) enediyne-associated epoxide hydrolase characterized to date. The identification of conserved residues at positions Y138, L187, A188, D189, P190, E191, H192, A194, F229, Q303, L304, and Q370 may represent critical sites for investigation of a robust catalytic toolkit for the production of enantiomerically pure diols from a wide variety of epoxide hydrolases and substrates. Finally, regioselectivity characterization of CynF and SghF has confirmed that CynF is a retaining epoxide hydrolase, and SghF is an inverting epoxide hydrolase, thereby expanding the inverting versus retaining paradigm for enediyne epoxide hydrolases.

References

(1) Van Loo, B.; Kingma, J.; Arand, M.; Wubbolts, M. G.; Janssen, D. B. *Appl. Environ. Microbiol.* **2006**, *72* (4), 2905.

(2) Widersten, M.; Gurell, A.; Lindberg, D. *Biochim. Biophys. Acta - Gen. Subj.* **2010**, *1800* (3), 316.

(3) Ollis, D. L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J.; Sussman, J. L.; Verschueren, K. H. G.; Goldman, A. **1990**, *5* (1989), 197.

(4) Argiriadi, M. a.; Morisseau, C.; Goodrow, M. H.; Dowdy, D. L.; Hammock, B. D.; Christianson, D. W. *J. Biol. Chem.* **2000**, *275* (20), 15265.

(5) Rink, R.; Janssen, D. B. *Biochemistry* **1998**, *37* (51), 18119.

(6) Rink, R.; Fennema, M.; Smids, M.; Dehmel, U.; Janssen, D. B. **1997**, *272* (23), 14650.

(7) Vicente, M. F.; Basilio, A.; Cabello, A.; Pelaez, F. *Clin. Microbial. Infect.* **2003**, *9*, 15.

(8) Khosla, C.; Gokhale, R. S.; Jacobsen, J. R.; Cane, D. E. *Annu. Rev. Biochem.* **1999**, *68*, 219.

(9) Walsh, C. T.; Chen, H.; Keating, T. a.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M. *Curr. Opin. Chem. Biol.* **2001**, *5* (5), 525.

(10) Liang, Z.-X. *Nat. Prod. Rep.* **2010**, *27*, 499.

(11) Lin, S.; Horsman, G. P.; Chen, Y.; Li, W.; Shen, B. *J. Am. Chem. Soc.* **2009**, *131* (45), 16410.

(12) Lin, S.; Horsman, G. P.; Shen, B. *Org. Lett.* **2010**, *12*, 3816.

(13) Horsman, G. P.; Lechner, A.; Ohnishi, Y.; Moore, B.; Shen, B. - *Biochemistry*. - American Chemical Society 2013, pp 5217–5224.

(14) Bornscheuer, U. T.; Kazlauskas, R. J.; Hulsman, G. W.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012**, *485*, 185.

(15) Griengl, H.; Schwab, H.; Fechter, M. *Trends Biotechnol.* **2000**, *18* (6), 252.

(16) Steinreiber, A.; Faber, K. *Curr. Opin. Biotechnol.* **2001**, *12* (6), 552.

(17) Rink, R.; Lutje Spelberg, J. H.; Pieters, R. J.; Kingma, J.; Nardini, M.; Kellogg, R. M.; Dijkstra, B. W.; Janssen, D. B. *J. Am. Chem. Soc.* **1999**, *121* (32), 7417.

(18) Lutz, S. *Curr. Opin. Biotechnol.* **2010**, *21* (6), 734.

(19) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2014**, *12* (1), 7.

(20) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. a; Huisman, G. W. *Nat. Biotechnol.* **2007**, *25* (3), 338.

(21) Jia, X.; Wang, Z.; Li, Z. *Tetrahedron: Asymmetry* **2008**, *19* (4), 407.

(22) Lutje Spelberg, J. H.; Rink, R.; Archelas, A.; Furstoss, R.; Janssen, D. B. *Adv. Synth. Catal.* **2002**, *344* (9), 980.

(23) Sun, Z.; Lonsdale, R.; Wu, L.; Li, G.; Li, A.; Wang, J.; Zhou, J.; Reetz, M. T. *ACS Catal.* **2016**, *6* (3), 1590.

(24) Krone, K. M.; Warias, R.; Ritter, C.; Li, A.; Acevedo-Rocha, C. G.; Reetz, M. T.; Belder, D. *J. Am. Chem. Soc.* **2016**, *138* (7), 2102.

(25) Littlechild, J. A. *Front. Bioeng. Biotechnol.* **2015**, *3* (October), 161.

(26) Nicolaou, K. C.; Huang, X.; Snyder, S. a; Rao, P. B.; Bella, M. **2002**, No. 5, 834.

(27) Pedragosa-Moreau, S.; Morisseau, C.; Baratti, J.; Archelas, A.; Furstoss, R. *Tetrahedron* **1997**, *53* (28), 9707.

(28) Szermerski, M.; Melesina, J.; Wichapong, K.; Löppenberg, M.; Jose, J.; Sippl, W.; Holl, R. *Bioorg. Med. Chem.* **2014**, *22* (3), 1016.

(29) Gerlt, J. a.; Allen, K. N.; Almo, S. C.; Armstrong, R. N.; Babbitt, P. C.; Cronan, J. E.; Dunaway-Mariano, D.; Imker, H. J.; Jacobson, M. P.; Minor, W.; Poulter, C. D.; Raushel, F. M.; Sali, A.; Shoichet, B. K.; Sweedler, J. V. *Biochemistry* **2011**, *50* (46), 9950.

(30) Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbit, P. C. *Plos Comput. Biol.* **2009**, *5* (12).

(31) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Arganiska, J.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Chavali, G.; Cibrian-Uhalte, E.; Da Silva, A.; De Giorgi, M.; Dogan, T.; Fazzini, F.; Gane, P.; Castro, L. G.; Garmiri, P.; Hatton-Ellis, E.; Hieta, R.; Huntley, R.; Legge, D.; Liu, W.; Luo, J.; Macdougall, A.; Mutowo, P.; Nightingale, A.; Orchard, S.; Pichler, K.; Poggioli, D.; Pundir, S.; Pureza, L.; Qi, G.; Rosanoff, S.; Saidi, R.; Sawford, T.; Shypitsyna, A.; Turner, E.; Volynkin, V.; Wardell, T.; Watkins, X.; Zellner, H.; Cowley, A.; Figueira, L.; Li, W.; McWilliam, H.; Lopez, R.; Xenarios, I.; Bougueleret, L.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Boeckmann, B.; Bolleman, J.; Boutet, E.; Breuza, L.; Casal-Casas, C.; De Castro, E.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Duvaud, S.; Estreicher, A.; Famiglietti, L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Jungo, F.; Keller, G.; Lara, V.; Lemercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T.; Nouspikel, N.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pozzato, M.; Pruess, M.; Rivoire, C.; Roechert, B.; Schneider, M.; Sigrist, C.; Sonesson, K.; Staehli, S.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Veuthey, A. L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Suzek, B. E.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L. S.; Yerramalla, M. S.; Zhang, J. *Nucleic Acids Res.* **2015**, *43* (D1), D204.

(32) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.;

Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Syst. Biol.* **2011**, *7* (1), 539.

(33) Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. *Nucleic Acids Res.* **2010**, *38* (SUPPL. 2), 695.

(34) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 213.

(35) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486.

(36) Cedrone, F.; Bhatnagar, T.; Baratti, J. C. *Biotechnol. Lett.* **2005**, *27* (23-24), 1921.

(37) Lohman, J. R.; Huang, S. X.; Horsman, G. P.; Dilfer, P. E.; Huang, T.; Wendt-Pienkowski, E.; Shen, B. *Mol Biosyst* **2013**, *9* (3), 478.

(38) Lacourciere, G. M.; Armstrong, R. N. *J. Am. Chem. Soc.* **1993**, *115* (Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.), 10466.

(39) Morisseau, C.; Du, G.; Newman, J. W.; Hammock, B. D. *Arch. Biochem. Biophys.* **1998**, *356* (2), 214.

(40) Bahl, C. D.; Hvorecny, K. L.; Morisseau, C.; Gerber, S. A.; Madden, D. R. **2016**.

(41) Clamp, J. R.; Hough, L. *Biochem. J.* **1966**, *101* (1), 120.

(42) Arroyo, M.; de la Mata, I.; Acebal, C.; Castillón, M. P. *Appl. Microbiol. Biotechnol.* **2003**, *60* (5), 507.

(43) Aboody, D.; Lev, B. R&D Productivity in the Chemical Industry, New York University, Stern School of Business, 2001.

(44) *Science* **1957**, *125* (3257), 1082.

Appendix 1. Additional Figures

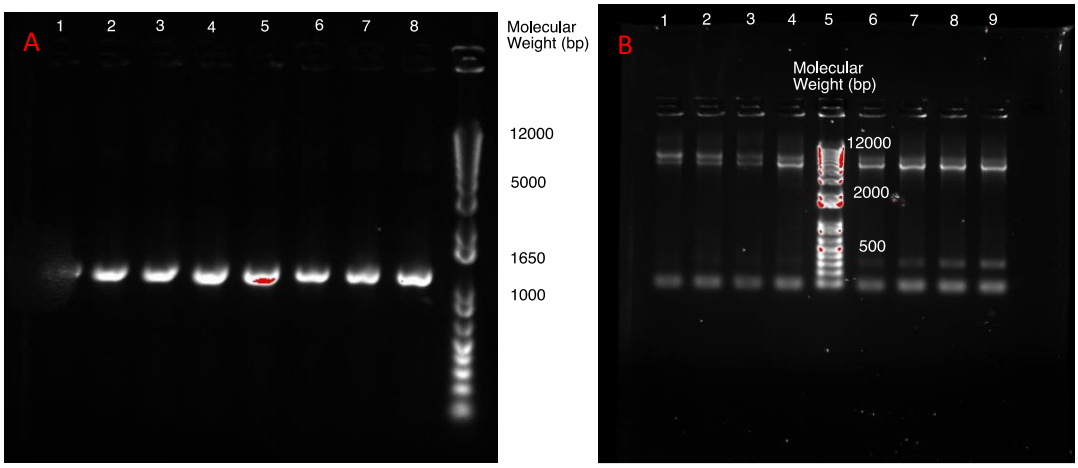


Figure A1. Preparation of SghF H364Q by overlap extension PCR. (A) PCR amplification of the portion of the gene upstream of the mutation, (B) PCR amplification of the portion of the gene downstream of the mutation. Homologous amplicons were pooled together and excised from the gel to increase DNA yield.

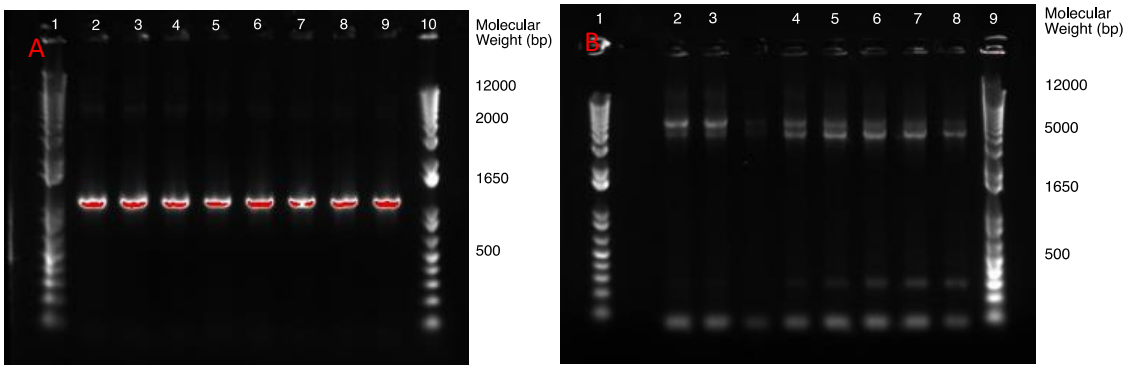


Figure A2. Preparation of SghF-H364A by overlap extension PCR. (A) PCR amplification of the portion of the gene upstream of the mutation, (B) PCR amplification of the portion of the gene downstream of the mutation. Homologous amplicons were pooled together and excised from the gel to increase DNA yield.

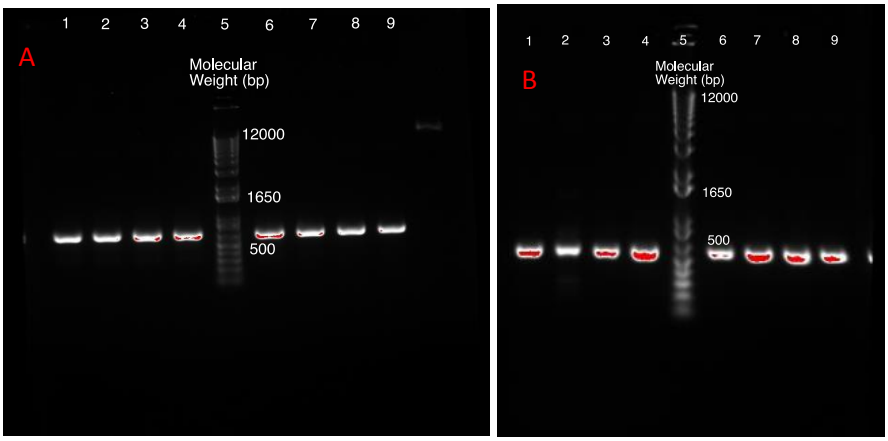


Figure A3. Preparation of CynF-D176N by overlap extension PCR. (A) PCR amplification of the portion of the gene upstream of the mutation, (B) PCR amplification of the portion of the gene downstream of the mutation. Homologous amplicons were pooled together and excised from the gel to increase DNA yield.

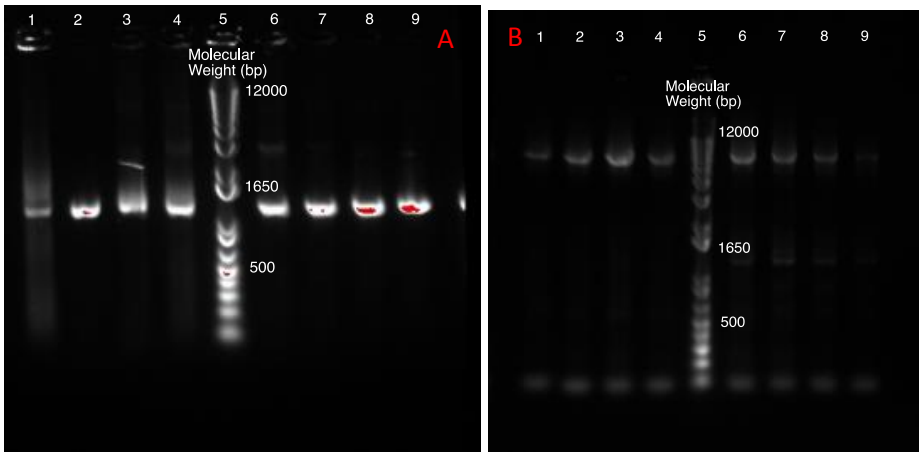


Figure A4. Preparation of CynF-H363Q by overlap extension PCR. (A) PCR amplification of the portion of the gene upstream of the mutation, (B) PCR amplification of the portion of the gene downstream of the mutation. Homologous amplicons were pooled together and excised from the gel to increase DNA yield.

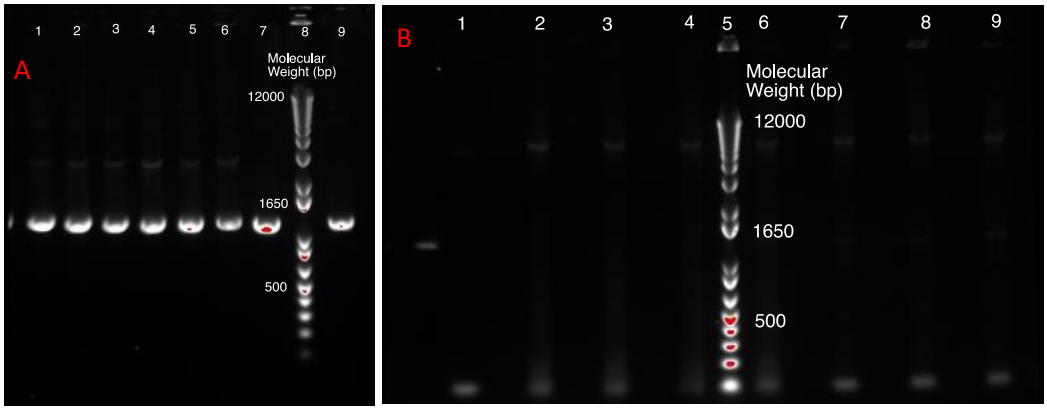


Figure A5. Preparation of CynF-H363A by overlap extension PCR. (A) PCR amplification of the portion of the gene upstream of the mutation, (B) PCR amplification of the portion of the gene downstream of the mutation. Homologous amplicons were pooled together and excised from the gel to increase DNA yield.

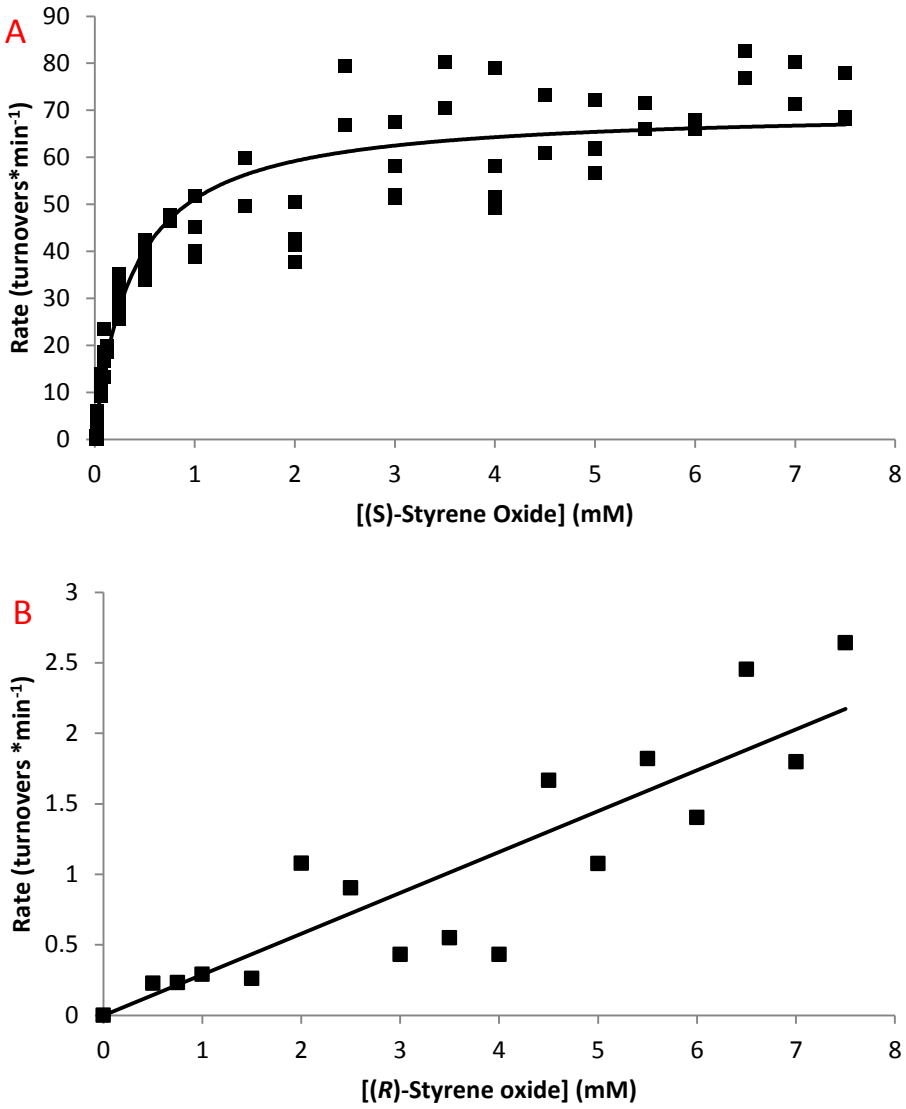


Figure A7. Kinetic analysis of *SghF*-catalyzed hydrolysis of (A) (*S*)-styrene oxide and (B) (*R*)-styrene oxide.

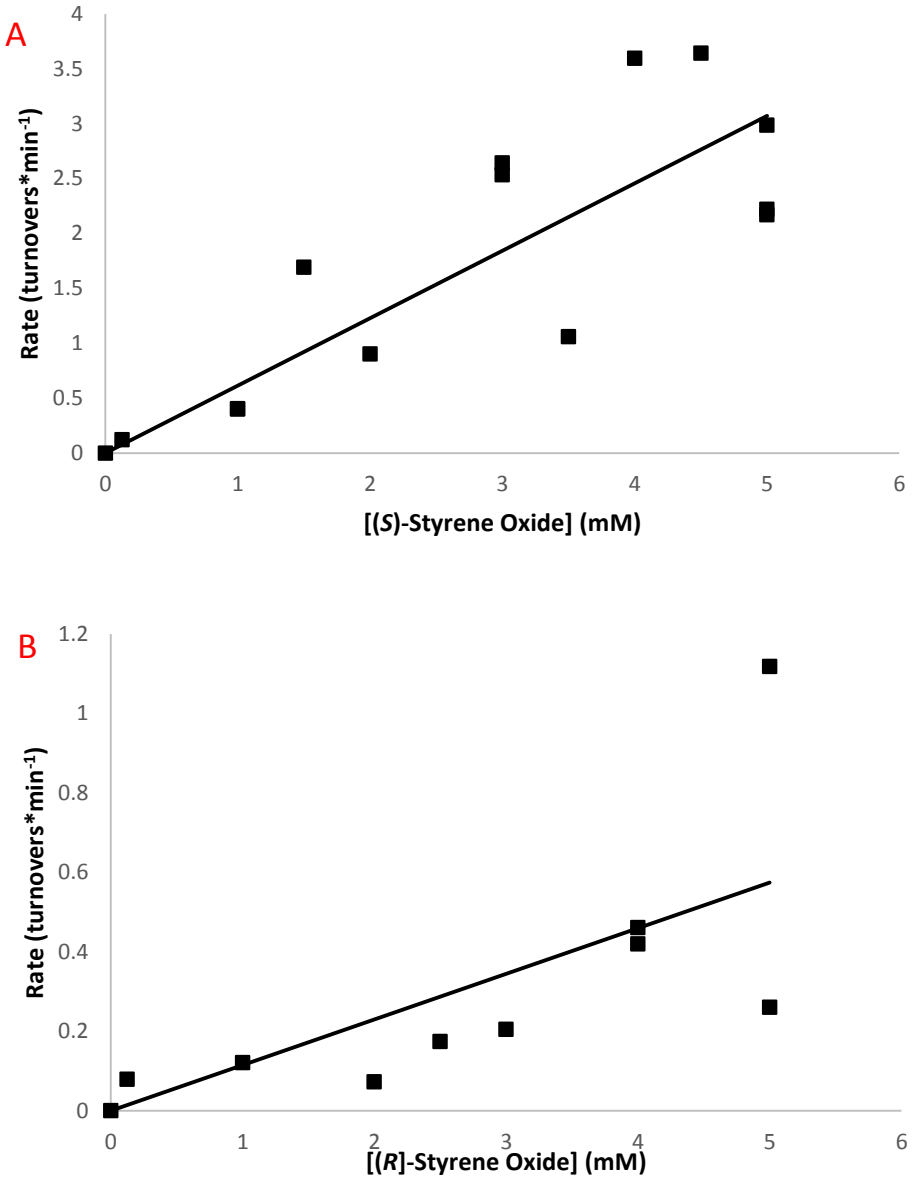


Figure A8. Kinetic analysis of *CynF*-catalyzed hydrolysis of (A) (*S*)-styrene oxide and (B) (*R*)-styrene oxide.

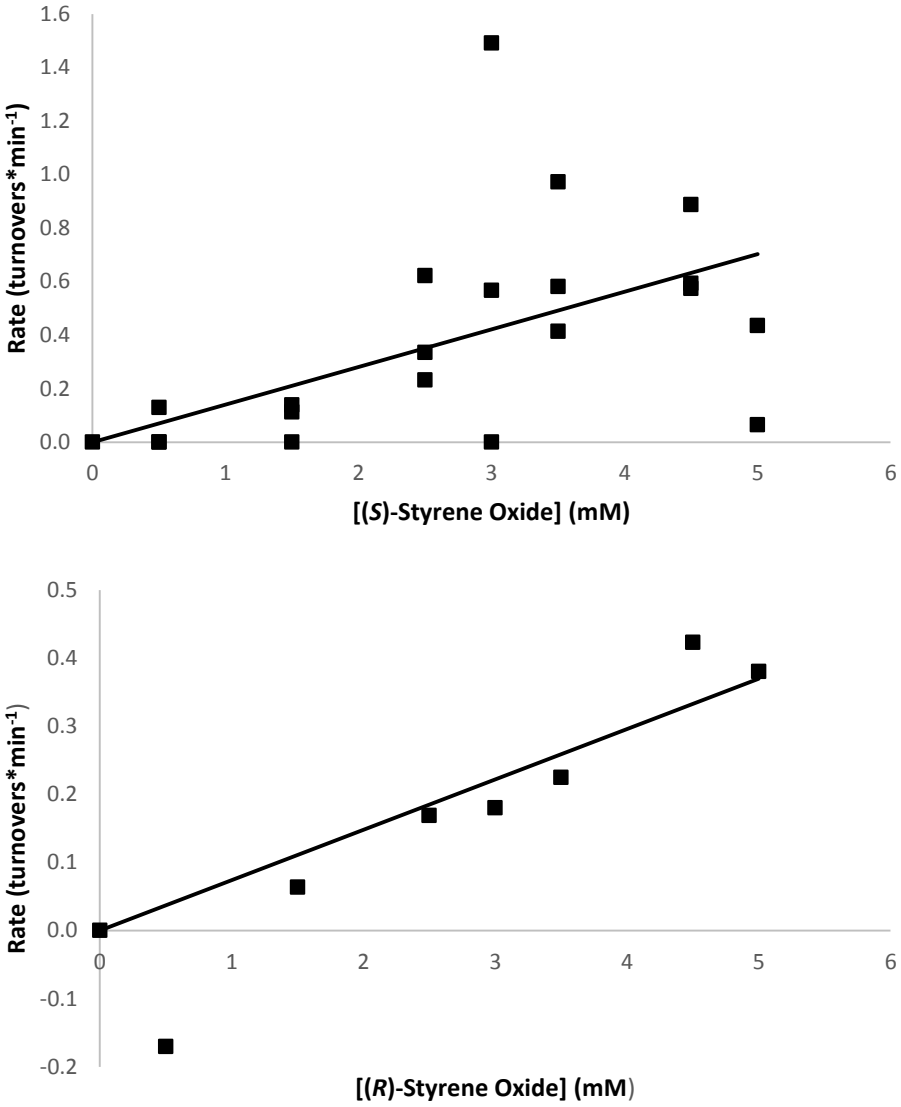


Figure A8. Kinetic analysis of SghF-D176N-catalyzed hydrolysis of (A) (*S*)-styrene oxide and (B) (*R*)-styrene oxide.

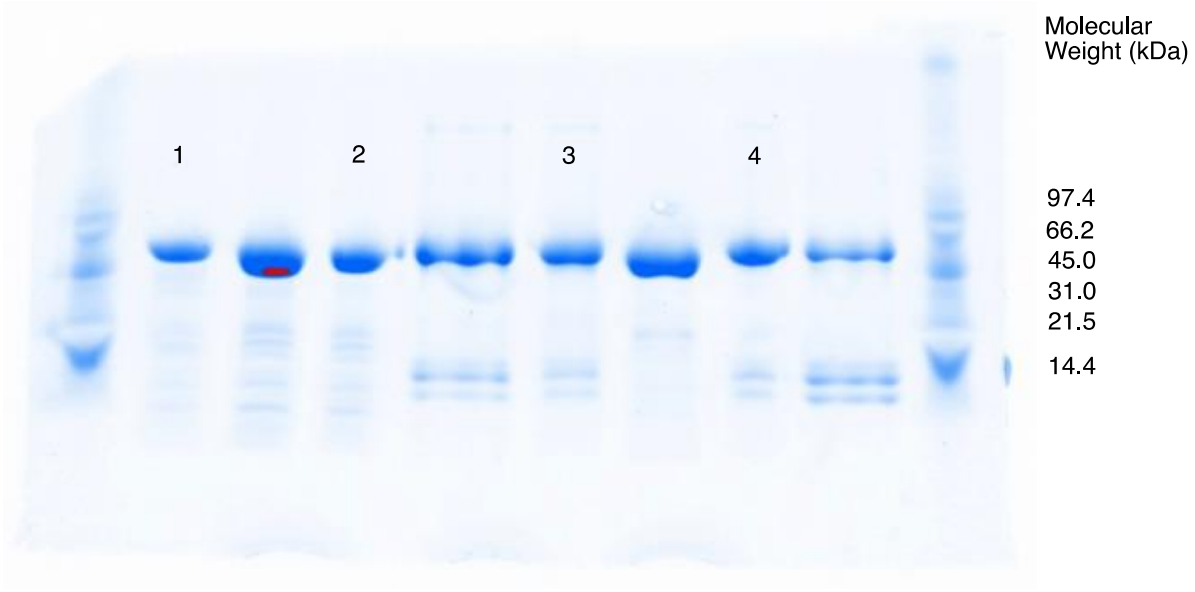


Figure A9. Overproduction and purification of (1) CynF-H363A, (2) SghF-D176N, (3) CynF-D175N, and (4) CynF to near homogeneity as judged by 12% SDS-PAGE.

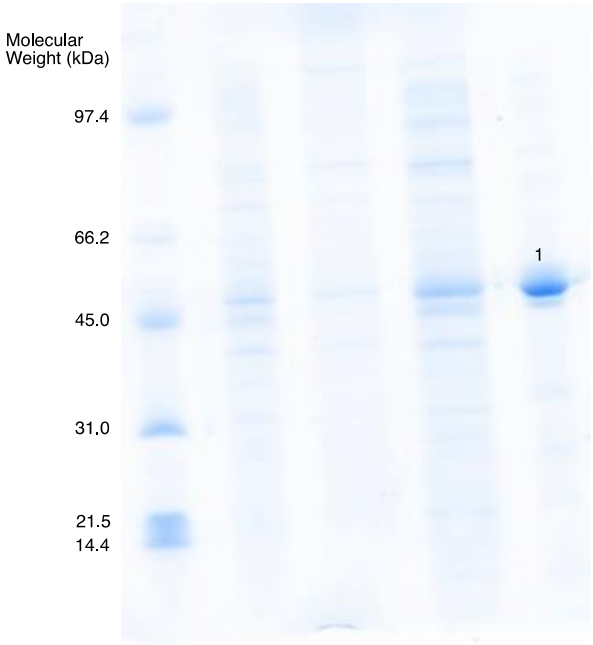


Figure A10. Overproduction and purification of (1) CynF-H363Q to near homogeneity as judged by 12% SDS-PAGE.

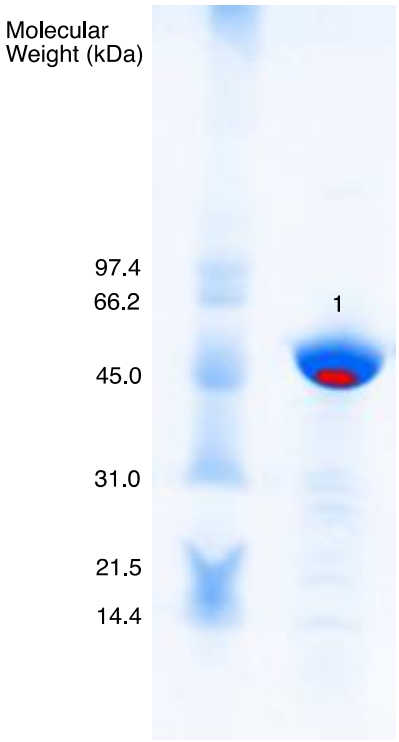


Figure A11. Overproduction and purification of (1) SghF-H364Q to near homogeneity as judged by 12% SDS-PAGE.

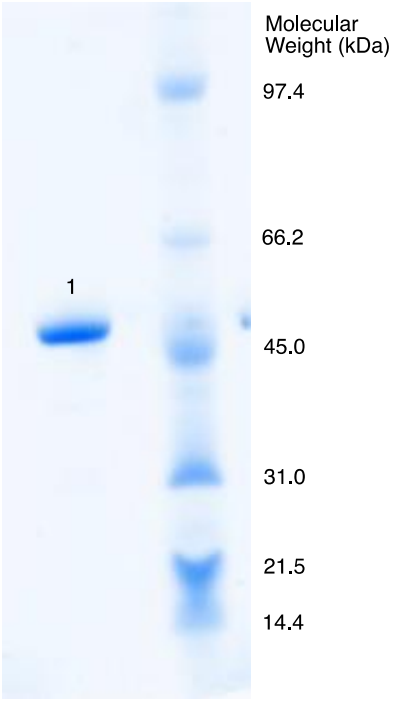


Figure A12. Overproduction and purification of (1) SghF-H364A to near homogeneity as judged by 12% SDS-PAGE.

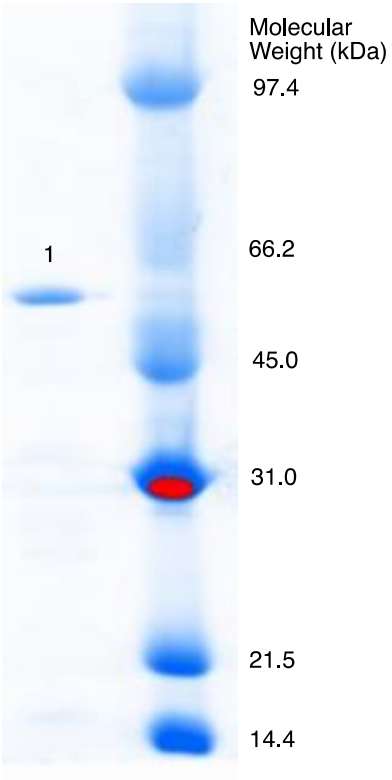
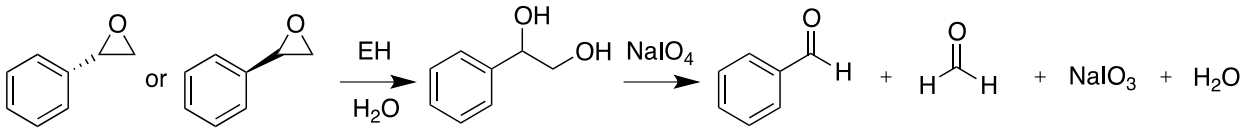


Figure A13. Overproduction and purification of (1) SghF to near homogeneity as judged by 12% SDS-PAGE.



Scheme A1. Kinetics reaction for monitoring the hydrolysis of styrene oxide by an epoxide hydrolase. Formation of benzaldehyde is monitored spectrophotometrically at 290nm.

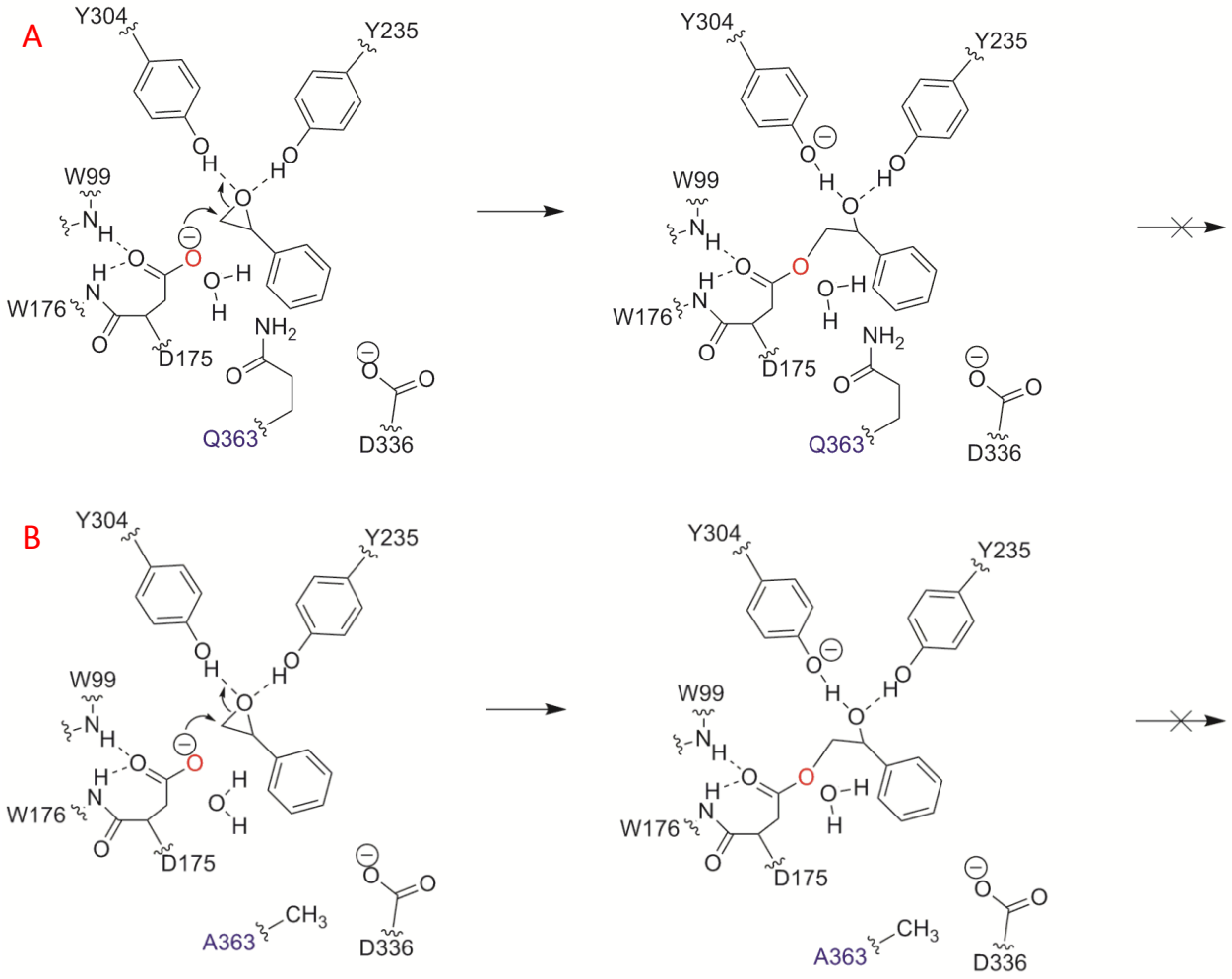


Figure A15. Proposed catalytic mechanism for hydrolysis of styrene oxide by (A) CynF-H363Q and (B) CynF-H363A.

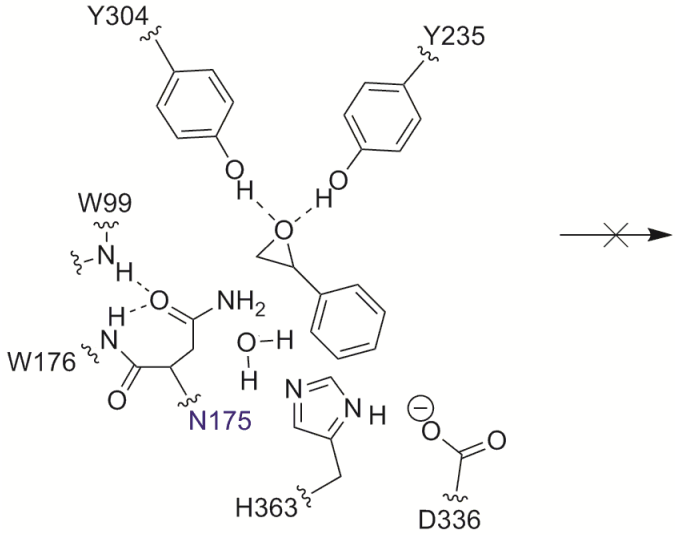


Figure A16. Propopsed mechanism for hydrolysis of styrene oxide by CynF-D175N

Table A1. Sequences used for EH bioinformatics analysis. Residue refers to the amino acid in position 236 (SghF numbering).

Residue	UniProt	Strain	Domain
W	A4X8F5	*Salinispora tropica* CNB-440	Bacteria
W	W2F1W8	*Microbispora* sp. ATCC PTA-5024	Bacteria
W	B1VRN6	*Streptomyces griseus* subsp. Griseus (strain JCM 4626 / NBRC 13350)	Bacteria
W	Q8GMH6	*Streptomyces globisporus*	Bacteria
W	D6AAC9	*Streptomyces ghanaensis* ATCC 14672	Bacteria
W	M0B613	*Natrialba asiatica* (strain ATCC 700177/ DSM 12278/ JCM9576/ FERM P-10747/ NBRC 102637/ 172P1)	Archaea
Y	L7EXA6	*Streptomyces turgidiscabies* Car8	Bacteria
Y	D3Q9C0	*Stackebrandtia nassauensis* (strain DSM 44728/ NRRL B-16338/ NBRC 102104/ LLR-40K-21)	Bacteria
Y	F3M694	*Paenibacillus* sp. HGF5	Bacteria
Y	F4F709	*Verrucosispora maris* (strain AB-18-032)	Bacteria
Y	D7BXM8	*Streptomyces bingchenggensis* (strain BCW-1)	Bacteria
Y	S4Y0k5	*Sorangium cellulosum* So0157-2	Bacteria